THE UNIVERSITY
of EDINBURGH

A CMOS SPAD-based Image Sensor *for* Single Photon Counting and Time of Flight Imaging

Neale A. W. Dutton

A thesis submitted for the degree of Doctor of Philosophy
THE UNIVERSITY *of* EDINBURGH
2015

To my wife Helena,

my father and stepmother,

and to the memory of my grandparents and my mother.

"Nothing in the world can take the place of persistence. Talent will not; nothing is more common than unsuccessful men with talent. Genius will not; unrewarded genius is almost a proverb. Education will not; the world is full of educated derelicts. Persistence and determination alone are omnipotent. The slogan 'Press On' has solved and always will solve the problems of the human race."

Calvin Coolidge

Supervised by:

Prof. Robert Henderson

The School of Engineering,
The University of Edinburgh,
The Kings Buildings,
Mayfield Road,
Edinburgh

Prof. Ian Underwood

The School of Engineering,
The University of Edinburgh,
The Kings Buildings,
Mayfield Road,
Edinburgh

Mr. Lindsay Grant

STMicroelectronics (R&D) Ltd.,
Imaging Division,
Pinkhill,
Edinburgh

Examined by:

Prof. Anthony Walton

The School of Engineering,
The Kings Buildings,
Mayfield Road,
Edinburgh

Dr David Stoppa

Fondazione Bruno Kessler,
via Santa Croce,
Trento,
Italy

On the 6th of April 2016

Abstract

The facility to capture the arrival of a single photon, is the fundamental limit to the detection of quantised electromagnetic radiation. An image sensor capable of capturing a picture with this ultimate optical and temporal precision is the pinnacle of photo-sensing. The creation of high spatial resolution, single photon sensitive, and time-resolved image sensors in complementary metal oxide semiconductor (CMOS) technology offers numerous benefits in a wide field of applications. These CMOS devices will be suitable to replace high sensitivity charge-coupled device (CCD) technology (electron-multiplied or electron bombarded) with significantly lower cost and comparable performance in low light or high speed scenarios. For example, with temporal resolution in the order of nano and picoseconds, detailed three-dimensional (3D) pictures can be formed by measuring the time of flight (TOF) of a light pulse. High frame rate imaging of single photons can yield new capabilities in super-resolution microscopy. Also, the imaging of quantum effects such as the entanglement of photons may be realised.

The goal of this research project is the development of such an image sensor by exploiting single photon avalanche diodes (SPAD) in advanced imaging-specific 130nm front side illuminated (FSI) CMOS technology. SPADs have three key combined advantages over other imaging technologies: single photon sensitivity, picosecond temporal resolution and the facility to be integrated in standard CMOS technology. Analogue techniques are employed to create an efficient and compact imager that is scalable to mega-pixel arrays. A SPAD-based image sensor is described with 320 by 240 pixels at a pitch of 8μm and an optical efficiency or fill-factor of 26.8%. Each pixel comprises a SPAD with a hybrid analogue counting and memory circuit that makes novel use of a low-power charge transfer amplifier. Global shutter single photon counting images are captured. These exhibit photon shot noise limited statistics with unprecedented low input-referred noise at an equivalent of 0.06 electrons.

The CMOS image sensor (CIS) trends of shrinking pixels, increasing array sizes, decreasing read noise, fast readout and oversampled image formation are projected towards the formation of binary single photon imagers or quanta image sensors (QIS). In a binary digital image capture mode, the image sensor offers a look-ahead to the properties and performance of future QISs with 20,000 binary frames per second readout with a bit error rate of 1.7×10^{-3}. The bit density, or cumulative binary intensity, against exposure performance of this image sensor is in the shape of the famous Hurter and Driffield densitometry curves of photographic film. Oversampled time-gated binary image capture is demonstrated, capturing 3D TOF images with 3.8cm precision in a 60cm range.

Lay Summary

The facility to capture the arrival of a single photon, is the fundamental limit to the detection of quantised electromagnetic radiation. An image sensor capable of capturing a picture with this ultimate optical and temporal precision is the pinnacle of photo-sensing. The creation of high spatial resolution, single photon sensitive, and time-resolved image sensors in complementary metal oxide semiconductor (CMOS) technology - the technology behind all modern electronic chips - offers numerous benefits in a wide field of applications. To detect single photons, this research employs a type of photo-diode that produces a measurable voltage pulse when it detects one photon called the single photon avalanche diode (SPAD).

SPADs have three key combined advantages over other imaging technologies: single photon sensitivity, picosecond temporal resolution and the facility to be integrated in standard CMOS technology. These CMOS devices will be suitable to replace high sensitivity charge-coupled device (CCD) technology, predominantly found in specialist images sensors for scientific imaging, with significantly lower cost and comparable performance in low light or high speed scenarios. For example, with temporal resolution in the order of nano and picoseconds, detailed three-dimensional (3D) pictures can be formed by measuring the time of flight (TOF) of a light pulse. High frame rate imaging of single photons can yield new capabilities in super-resolution microscopy. Also, the imaging of quantum effects such as the entanglement of photons may be realised.

Recent image sensors, as found in your mobile phone camera to DSLRs to broadcast cameras are built using CMOS technology. CMOS image sensor (CIS) trends of shrinking pixels, increasing array sizes, decreasing read noise, fast readout and oversampled image formation are projected towards the formation of binary single photon imagers or Quanta image sensors (QIS). In a binary digital image capture mode, the image sensor presented in this research offers a look-ahead to the properties and performance of future QISs with 20,000 binary frames per second readout with a low error rates. The bit density, or cumulative binary intensity, against exposure performance of this image sensor is in the shape of the famous Hurter and Driffield densitometry curves of photographic film. The image sensor presented in this thesis is capable of capturing single photon images with an unprecedented low noise level.

Declaration of Originality

I hereby declare that the research recorded in this thesis (excluding the exceptions stated below) and the thesis itself originated with and was composed entirely by myself.

The CMOS pixel structures outlined in Chapter 4 and Appendix 1 were conceived by Prof. Robert Henderson of The University of Edinburgh, then modified and adapted through joint effort with myself.

The design and implementation of the CMOS SPAD image sensor, described in Chapter 5, was assisted by both Prof. Robert Henderson and Luca Parmesan of The University of Edinburgh.

Other people who have contributed to this work are acknowledged and / or referenced appropriately within.

This work has not been submitted for any other degree or professional qualification except as specified.

Neale A.W. Dutton

Acknowledgements

My first thanks are to my supervisor Prof. Robert Henderson whose ideas, support, expert advice, and enthusiasm made this an incredible doctoral experience and a very worthwhile four years. There are few like him, with such drive, perseverance, energy, good humour and always many steps ahead of everyone else. Congratulations Robert on your chair in Electronic Imaging, a well-deserved promotion. I am grateful to have been mentored in the first year of my PhD by Dr David Renshaw; he provided support, enthusiasm and straight-talking advice at the start of the project. Over the course of my PhD I have received useful advice, direction and inspiration from Prof. Ian Underwood to whom I am thankful.

I am truly grateful to my industrial supervisor Dr Lindsay Grant for his guidance and mentorship over the last four years. I am indebted to him for his support and advice, and for the many opportunities he has helped to create in my career. Thanks go to Dr Andrew Holmes for his frank advice, humour, and many discussions and design reviews. Dr Sara Pellegrini provided scientific advice, support, and encouragement. Joint thanks are due to both Drew and Sara for their kind donation of time of flight measurement equipment, and to Dr Bruce Rae for many technical discussions and his support throughout this project. My thanks and gratitude are also due to many further people at ST, in particular: Stuart McLeod, Kevin Moore, Dave Poyner, Dr Graeme Storm, Jeff Raynor (for cake and discussions), Dr Robert Nicol, Brent Hearn, Dave Grant, Dave Lee, Steve East, Jim Gallagher, Johannes Vergote, and Gary McGillvary for interviewing me and taking me on as a 3rd year summer student in 2009! Thanks also to Nicky Grant for helping solve many a CAD tools problem.

To all three of my fellow ST funded PhD students: good luck in your thesis write ups! Laurence Stark, you never cease to surprise me with your originality. Salvatore Gnecchi, I have really enjoyed our collaboration on the folded flash TDC and the optimisation of SPAD arrays, good luck in your future career. Luca Parmesan, I'm indebted to you for your great work on my PhD image sensor and I'm glad we could work together on yours, I've always appreciated your can do/get it done attitude. Aravind Venugopalan, good luck in your write up and in your new adventure as a father! Tarek Al Abbas, it's been a pleasure working and collaborating with you over the last two years, and I look forward to continuing in the next two, thanks also for proof-reading this thesis. Thanks to Francescopaolo Mattioli Della Rocca for proof-reading this thesis.

I am grateful to Dr Richard Walker for his support and attention particularly at the start of my PhD, and I appreciate our continued friendship. Thanks to Dr Istvan Gyongy for his support and an enjoyable collaboration under this research project. Many thanks to my University colleagues for their time, guidance, support and friendship in particular: Dr Katherine Cameron, Dr Nikola Krstajic, Dr Eric Webster and Dr David Tyndall. Thank you to Dr John Jeffrey for good advice and humour, and particularly for his part in keeping the ST RCA going. Also I am grateful to the late Susan Kivlin, for her assistance with this project.

I am glad to have so many great friends and family who have supported and distracted me throughout my PhD. Good luck to Dave, Chantal, Jake and my sister in-law Christina in their own research adventures. I am truly thankful for the advice, love, fun and support of Jane and Andy, Hannah and Emma. Many thanks to my Professor parents-in-law Sheila and Mike for sound academic advice, wisdom and encouragement. To my siblings, Fiona and Kiron for never stopping to be great fun, loving and supporting. To Kirsteen for her love, generosity, warmth and kindness and never ceasing to keeping me stocked in snacks! I am so lucky to have such a role model as my father, Prof. Gordon Dutton, whose constant advice, teaching and endless support has been fundamental to the success of this research project, my career and my education. Alongside the role models of my late mother and grandfather Arthur, I am eternally grateful to them all for giving me my persistence, drive and enthusiasm.

My final thanks and eternal gratitude are, of course, to my wife and best friend Helena who has endured four years of lost evenings, weekends and holidays to my PhD. I am looking forward to spending much more time with her after this thesis submission. Helena, thank you for keeping me on track and organised, especially when I was reluctant to be either of those things. Your remarkable patience, attention and never ending pride and love is incredible. I am so proud of you, and being able to call you my wife is a privilege that bears no equal.

Publications

Published Conference Papers

[1] Neale A. W. Dutton, Lindsay A. Grant, and Robert K. Henderson, "9.8μm SPAD-based Analogue Single Photon Counting Pixel with Bias Controlled Sensitivity" in *International Image Sensor Workshop*, June 2013.

[2] Neale A. W. Dutton, Luca Parmesan, Andrew J. Holmes, Lindsay A. Grant, and Robert K. Henderson, "320x240 Oversampled Digital Single Photon Counting Image Sensor," in *IEEE VLSI Symposium*, June 2014.

[3] Salvatore Gnecchi, Neale A. W. Dutton, Johannes Vergote, Lindsay A. Grant, David Lee, Sara Pellegrini, Bruce R. Rae, and Robert K. Henderson, "Multiple-event Direct to Histogram TDC in 65nm FPGA Technology," in *IEEE 10th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME)*, July 2014

[4] Luca Parmesan, Neale A. W. Dutton, Neil J. Calder, Andrew J. Holmes, Lindsay A. Grant, and Robert K. Henderson, "A 9.8μm Sample and Hold Time to Amplitude Converter CMOS SPAD Pixel," in *IEEE ESSDERC*, September 2014.

[5] Neale A. W. Dutton, Salvatore Gnecchi, Luca Parmesan, Andrew J. Holmes, Bruce R. Rae, Lindsay A. Grant, and Robert K. Henderson, "A Time Correlated Single Photon Counting Sensor with 14GS/s Histogramming Time to Digital Converter," in *ISSCC, IEEE International Solid-State Circuits Conference Digest of Technical Papers, Paper 11.5,* February 2015

[6] Neale A. W. Dutton, Luca Parmesan, Salvatore Gnecchi, Istvan Gyongy, Neil J. Calder, Bruce R. Rae, Lindsay A. Grant, and Robert K. Henderson, "Oversampled ITOF Imaging Techniques using SPAD-based Quanta Image Sensors," in *International Image Sensor Workshop*, June 2015.

[7] Luca Parmesan, Neale A. W. Dutton, Neil J. Calder, Lindsay A. Grant, and Robert K. Henderson, "A 256x256 SPAD array with in-pixel Time to Amplitude Conversion for Fluorescence Lifetime Imaging Microscopy," in *International Image Sensor Workshop*, 2015.

[8] Istvan Gyongy, Neale A. W. Dutton, Luca Parmesan, Amy Davies, Rebecca Saleeb, Rory Duncan, Colin Rickman, Paul Dalgarno, and Robert K. Henderson, "Bit-plane Processing Techniques for Low-Light, High Speed Imaging with a SPAD-based QIS," in *International Image Sensor Workshop*, June 2015.

[9] Oscar Almer, Neale A. W. Dutton, Tarek Al Abbas, Salvatore Gnecchi, and Robert K. Henderson "4-PAM Visible Light Communications with a XOR-tree Digital Silicon Photomultiplier" in *IEEE Photonics Society, Summer Topicals*, July 2015.

[10] Salvatore Gnecchi, Neale A. W. Dutton, Luca Parmesan, Bruce R. Rae, Sara Pellegrini, Lindsay A. Grant, and Robert K. Henderson, "Gphotons/s TCSPC with XOR Digital Silicon Photomultipliers," in *Single Photon Workshop*, November 2015.

Accepted Conference Papers

[11] Salvatore Gnecchi, Neale A. W. Dutton, Luca Parmesan, Bruce R. Rae, Sara Pellegrini, Stuart J. McLeod, Lindsay A. Grant, and Robert K. Henderson, "A Comparative Analysis between OR-based and XOR-based Digital Silicon Photomultipliers for PET," in *IEEE Nuclear Science Symposium*, 2015.

[12] Oscar Almer, Dobroslav Tsonev, Neale A. W. Dutton, Tarek Al Abbas, Stefan Videv, Salvatore Gnecchi, Harald Haas, and Robert K. Henderson, "A SPAD-based Visible Light Communications Receiver Employing Higher Order Modulation," in *IEEE Globecom, 2015*.

Conference Abstracts

[13] Simon P. Poland, Nikola Krstajic, James Monypenny, Simao Coelho, David Tyndall, Richard J. Walker, Viviane Devauges, Justin Richardson, Neale A. W. Dutton, David Day-Uei Li, Klaus Suhling, Maddy Parsons, Tony Ng, Robert K. Henderson and Simon M. Ameer-Beg "A Time-Resolved Multifocal Multiphoton Microscope for High Speed Fret Imaging In-Vivo," in *Focus on Microscopy,* April 2014.

[14] Robert K. Henderson, Nik Krstajic, Neale A. W. Dutton, and Richard J. Walker, "CMOS Single Photon Image Sensors For Microscopy," in *Microscience and Microscopy Congress (MMC)*, June 2014.

[15] Simon P. Poland, Nikola Krstajic, James Monypenny, Simao Coelho, David Tyndall, Richard J. Walker, Viviane Devauges, Justin Richardson, Neale A. W. Dutton, David Day-Uei Li, Klaus Suhling, Maddy Parsons, Tony Ng, Robert K. Henderson and Simon M. Ameer-Beg "A Time-Resolved Multifocal Multiphoton Microscope for High Speed Fret Imaging In-Vivo," in *Microscience and Microscopy Congress (MMC),* June 2014.

Published Peer-Reviewed Journal Papers

[16] S. P. Poland, N. Krstajić, J. Monypenny, S. Coelho, D. Tyndall, R. J. Walker, V. Devauges, J. Richardson, Neale A.W. Dutton, P. Barber, D. D.-U. Li, K. Suhling, T. Ng, R. K. Henderson, and S. M. Ameer-Beg, "A high speed multifocal multiphoton fluorescence lifetime imaging microscope for live-cell FRET imaging," Biomedical Optics Express, vol. 6, no. 2, p. 277, Jan. 2015.

[17] Neale A.W. Dutton, Istvan Gyongy, Luca Parmesan, Salvatore Gnecchi, Neil J. Calder, Bruce R. Rae, Sara Pellegrini, Lindsay A. Grant and Robert K.Henderson "A SPAD-based QVGA Image Sensor for Single Photon Counting and Quanta Imaging", IEEE Transactions of Electron Devices Special Issue on Solid-State Image Sensors, January 2016.

[18] Salvatore Gnecchi, Neale A. W. Dutton, Luca Parmesan, Bruce R. Rae, Sara Pellegrini, Stuart J. McLeod, Lindsay A. Grant, and Robert K.Henderson "Digital Silicon Photomultipliers With OR/XOR Pulse Combining Techniques", IEEE Transactions of Electron Devices, March 2016

[19] Salvatore Gnecchi, Neale A. W. Dutton, Luca Parmesan, Lindsay A. Grant, and Robert K.Henderson "Analysis of Photon Detection Efficiency and Dynamic Range in SPAD based Visible Light Receivers", IEEE Journal of Lightwave Technology, April 2016

Accepted Peer-Reviewed Journal Paper

[20] Neale A. W. Dutton, Istvan Gyongy, Luca Parmesan, and Robert K. Henderson "Single Photon Counting Performance and Noise Analysis of CMOS SPAD-based Image Sensors" Journal of Sensors, Accepted for Publication January 2016.

Patents

Granted or Filed Patents

[21] Neale A. W. Dutton "Real-Time Processing Method and System," USPO Patent US 20120162632 A1, 28 June 2012.

[22] Neale A. W. Dutton, "Error Correction in Thermometer Codes," USPO Patent US 8432304 B2, 30 April 2013.

[23] Neale. A. W. Dutton, "Circuit for combining signals," USPO Patent US 9007118 B2, 14 April 2015.

[24] Neale A. W. Dutton, Robert K. Henderson "Improvements in time of flight pixel circuits," European Patent Office Patent 2728373 A1, USPO Patent US20140124652, 7 May 2014.

[25] Neale A. W. Dutton, Robert K. Henderson "Pixel circuit with capacitor discharge indicative of number of events," USPO Patent 20140124653 A1, 8 May 2014.

[26] Neale A. W. Dutton, Salvatore Gnecchi, Robert K. Henderson "Time to digital converter and applications thereof," USPO Patent US20150041625 A1, 12 February 2015.

Acronyms

APD	Avalanche Photo Diode
AQ	Active Quench
AQAR	Active Quench, Active Recharge
AQPR	Active Quench, Passive Recharge
ASiPM	Analogue Silicon Photo Multiplier
AR	Active Recharge
BG	Background (TOF Ambient)
BSI	Back Side Illuminated
CAPD	Current Assisted Photo Demodulator
CCD	Charge Coupled Device
CDS	Correlated Double Sampling
CIS	CMOS Image Sensor
CMOS	Complementary Metal Oxide Semiconductor
CPGA	Ceramic Pin Grid Array IC Package
CSS	CMOS Sensors and Systems Research Group
CTA	Charge Transfer Amplifier
DDS	Delta Differential Sampling
DNW	Deep N-well (in some cases Deep Retrograde N-Well)
DRM	Design Rule Manual
DSP	Digital Signal Processing
DSiPM	Digital Silicon Photomultiplier
DTC	Digital to Time Converter
DTOF	Direct Time of Flight
EBCCD	Electron Bombarded Charge Coupled Device
EMCCD	Electron Multiplication Charge Coupled Device
EBCMOS	Electron Bombarded CMOS Image Sensor
FLIM	Fluorescence Lifetime Imaging Microscopy
FOM	Figure of Merit
FPS	Frames Per Second
FRET	Forster Resonance Energy Transfer
FSI	Front Side Illuminated
FSM	Finite State Machine
GAPD	Geiger mode Avalanche Photo Diode (see SPAD)
GPU	Graphical Processing Unit
GRO	Gated Ring Oscillator
HCI	Human Computer Interface
HFPN	Horizontal Fixed Pattern Noise
IC	Integrated Circuit
ISP	Image Signal Processing

ITOF	Indirect Time of Flight
LDO	Low Drop Out Regulator
LOFIC	Lateral Overflow Integration Capacitor
MOM	Metal Oxide Metal Capacitor
MTF	Modulation Transfer Function
NI	National Instruments
NIM	Nuclear Instrumentation Module
NW	N-Well
Op-Amp	Operational Amplifier
PALM	Photo-activated Localisation Microscopy
PDK	Process Design Kit
PMD	Photonic Mixing Device
PPFPN	Pixel to Pixel Fixed Pattern Noise
PQ	Passive Quench
PQAR	Passive Quench, Active Recharge
PQPR	Passive Quench, Passive Recharge
PR	Passive Recharge
PW	P-well
QVGA	Quarter Video Graphics Array
SCS	Switched Current Source
SFDR	Spurious Free Dynamic Range
SiPM	Silicon Photo-Multiplier
SNR	Signal to Noise Ratio
SPAD	Single Photon Avalanche Diode
SRS	Stanford Research Systems
STI	Shallow Trench Isolation
STORM	Stochastic Optical Reconstruction Microscopy
TAC	Time to Amplitude Converter
TCSPC	Time Correlated Single Photon Counting
TDC	Time to Digital Converter
TSV	Through Silicon Via
USPO	United States Patent Office
VFPN	Vertical Fixed Pattern Noise
VLC	Visible Light Communications

Preface

This thesis describes and documents research undertaken through the joint research and collaboration agreement between STMicroelectronics' (ST) Imaging Division and Prof. Robert Henderson's 'CMOS Sensors and Systems' (CSS) research group at the University of Edinburgh. It builds on research and development work and substantial technological progress made by the PhD students who preceded me namely: Dr Bruce Rae, Dr Justin Richardson, Dr Richard Walker, Dr David Tyndall and Dr Eric Webster all under the attentive and invaluable supervision of Prof. Robert Henderson.

Under this joint venture with ST, access was provided to simulation tools, and crucially to fabrication of our CMOS integrated circuits (IC) without which our PhD projects would have been impossible. This is simply an opportunity that very few PhD students get. To ST, having people working on proof-of-concept ideas is immensely valuable in discovering what works, what works well and vitally what doesn't work.

One may learn something if one's design works, but one learns a tremendous amount through making mistakes and seeing a chip failing in ways one cannot conceive at the design stage. I am grateful to Robert and the ST team (Lindsay Grant, Sara Pellegrini, Bruce Rae and Andrew Holmes in particular) for allowing me to make mistakes, minor and disastrous, and allowing me to pursue my own directions at my own pace.

Whilst attending conferences, I am grateful to have met, engaged in conversations and discussions, debated, and received encouragement and technical feedback from many high-profile figures in the imaging field: (in no particular order) Professor Eric Fossum, Professor Nobukazu Teranishi-san, Dr Howard Rhodes and Dr Johannes Solhusvik, Dr Boyd Fowler, Prof Makato Ikeda-san, Prof Edoardo Charbon, Dr Cristiano Niclass, Dr David Stoppa, Dr Lucio Pancheri, and Dr Matteo Perenzoni. I was particularly honoured to meet and converse with the inventor of the charge coupled device (CCD) image sensor Dr Michael Tompsett (although he was not recognised in the citation for the Nobel Prize for the CCD image sensor) and Peter Noble, the inventor of the MOS active pixel sensor (APS) circuit. Photographs of these meetings take pride of place in Figures P.1 and P.2.

A doctorate is what you make of it, and I have endeavoured to make the most of mine. I encourage my current PhD colleagues and all future PhD students to do the same.

Neale A.W. Dutton, July 2015

Figure P.1. The author meeting Michael Tompsett, the inventor of the CCD image sensor at the International Image Sensors Workshop, in Utah, USA, 2013.

Figure P.2. The author meeting Peter Noble, the inventor of the MOS active pixel sensor at the International Image Sensors Workshop, in Vaals, the Netherlands, 2015.

Contents

Figures

Tables

1. Introduction

This thesis explores the creation of a high spatial resolution complementary metal oxide semiconductor (CMOS) image sensor based on single photon avalanche diode (SPAD) technology. SPADs have two distinct characteristics: single photon sensitivity and picosecond temporal resolution. Combining both of these, in a high resolution imager, yields a sensor operating at the fundamental limit of the detection of quantised electromagnetic radiation. With the capability of capturing a picture with the ultimate sensitivity in both optical and temporal domains.

1.1. Time for Temporal Imaging

This introductory section looks at the evolution of image sensors towards single photon counting and time-domain or 'temporal' imaging. The timeline of image sensors is marked with a number of significant milestones. Two inventions paved the way for CMOS imaging: Gene Weckler first described the photon integration mode of a photo-diode controlled by a pulsed reset FET and a readout FET in 1965 [27], and Peter Noble conceived a 5T source follower readout circuit in 1968 that was later renamed the active pixel sensor (APS) [28]. Yet, before CMOS imaging became mainstream, CCD technology dominated. In 1971, Michael Tompsett filed the first patent on a charge transfer imaging pixel or charge-coupled device (CCD) and soon after took the first colour 'digital' photograph of his wife with a CCD image sensor he co-invented with Smith and Boyle in Bell Labs in 1969 [29], ushering in the replacement of photographic film with silicon. By the 1990's CCD technology had matured and was the market leader in image sensors. In 1993, Eric Fossum correctly predicted that CCDs would lose market share dominance, becoming 'extinct', to be replaced by the CMOS APS in the majority of image sensor applications. Due to the lower-power and higher frame rate achievable with CIS, these were adopted into consumer and mobile electronic devices and overtook the CCD. Teranishi's invention of the isolated 'PIN' photo-diode provided many advantages to the conventional photo-diode based 3T circuit (e.g. kT/C noise mitigation). The combination of the APS circuit and PIN photo-diode became the standard approach CIS for over 15 years and is now a mature technology in billions of cameras. The main incremental developments in the field of CIS have been shrinking the pixel pitch down to a single micron, increasing array size, increasing the conversion gain of the pixel, and lowering read noise whilst increasing frame rate.

In parallel, from the 1950's to today, time-domain optical sensing developed using vacuum-tube based photo-multiplier tubes, discrete photo-diodes, or avalanche photo-diodes (APD) combined with external analogue

signal processing and high speed analogue to digital converters (ADC) [30]. Three key advances permitted time-resolved sensing and time-domain image capture to move from discrete diode, and accompanying external components, to being fully-integrated on a CMOS chip. First, in 1995, Spirig and Seitz proposed the first time resolved imaging pixel using a CCD-style photo-gate, moving charge synchronous to a modulated light source and ushering in the beginning of time-domain image sensors [31]. Aull's work in creating arrays of time converters for laser detection and ranging (LIDAR) image sensors, were the first imagers based on the time-correlated single photon counting (TCSPC) approach [32]. And third, in 2002 Rochas and Popovic implemented the first CMOS SPAD fully integrated with readout electronics [33].

In CISs, lowering the read noise would eventually allow the detection and wide-field imaging of single photons without expensive external components (such as cooling and image intensifiers). Further combining this with time-detection per pixel, allowed the measurement and high resolution imaging of the times of arrival of individual photons. Such time-resolved single photon imaging is an ideal research topic. In comparison to CIS, it is a relatively unexplored field with scope for advance. The single photon avalanche diode (SPAD) is an appropriate and suitable device to investigate, understand and develop this imaging modality for a variety of reasons. Notably, the detector has single photon sensitivity, picosecond temporal resolution, and is amenable to implementation with CMOS logic. Most importantly, it can be reliably manufactured in a CMOS imaging process with low mismatch and low dark noise.

The capability of SPAD-based sensors goes beyond these traditional CMOS and CCD based technologies, capturing single photon images with picosecond temporal accuracy. It facilitates time of flight (TOF) based applications such as 3D vision and LIDAR [5], [6], [34]–[39], as well as time correlated single photon counting (TCSPC) applications such as time-domain fluorescence lifetime imaging microscopy (FLIM) [4], [7], [40], Förster resonance energy transfer (FRET) FLIM [15], [16], TOF positron emission tomography (PET) [41], and imaging of ultrafast physical processes such as light in flight [42]. Moreover, it can bring benefits to high frame rate single photon counting applications such as stochastic optical reconstruction microscopy (STORM) and photo-activated localisation microscopy (PALM) [8].

Despite the single photon sensitivity of individual SPAD detectors, it has until now been challenging to assemble SPAD image sensors with sufficient spatial resolution and optical fill-factor to realize low-light CMOS image sensors. Consumer and scientific CISs have made, and are continually making, progress towards single photon counting regime by means of increased pixel conversion gain, low temporal noise and multiple sampling readout [43]. However, the condition of lower than 0.15e- readout noise for true single photon counting imaging has yet to be met.

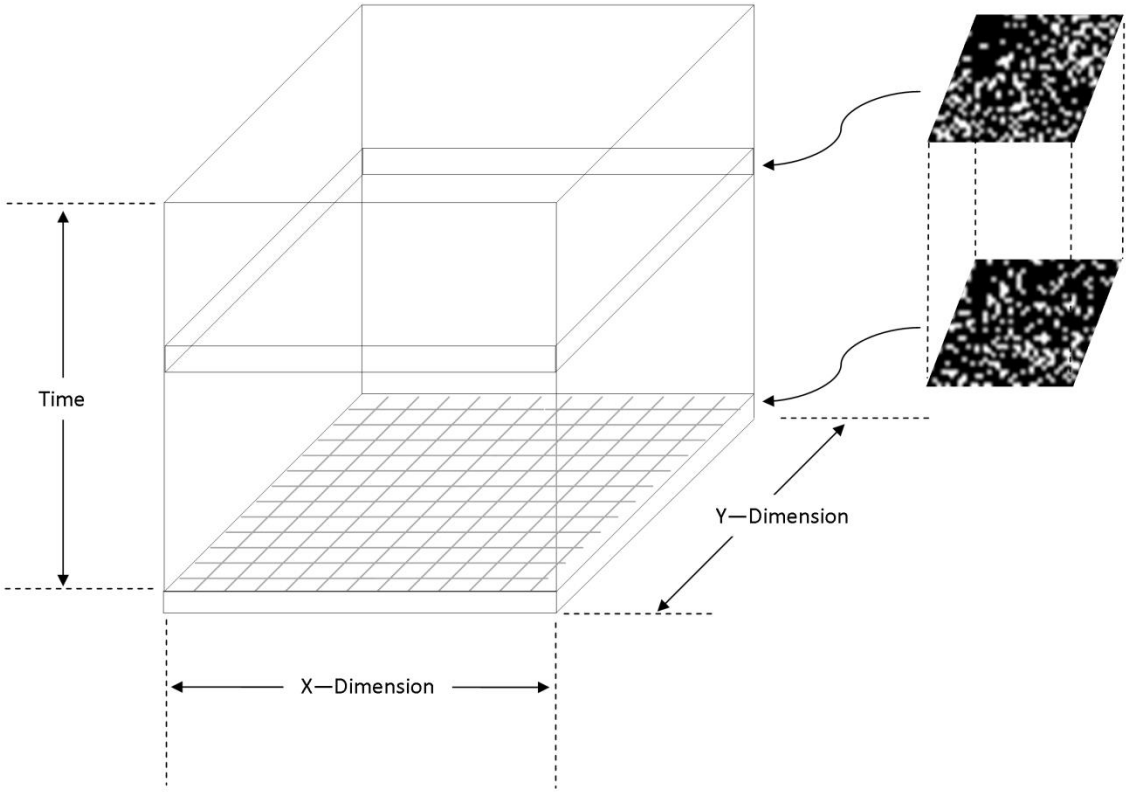


Figure 1.1. An illustration of spatial and temporal oversampling of single photon images demonstrating the Quanta Image Sensor (QIS) concept proposed by Fossum [44].

Prof Eric Fossum projected the CIS trends of pixel shrink, megapixel arrays, low read noise, fast readout, and highly oversampled image formation towards the Quanta Image Sensor (QIS) in 2005 [45], [46]. The QIS is an imaging array of photo- sensitive sites with single bit output, each site referred to as 'a jot'. The pixel concept of the QIS (a "jot") demands a nano-scale single photon sensitive device exhibiting a binary state. Figure 1.1 illustrates binary field images (or jot bit planes) oversampled either spatially or temporally (or both). The output of this operation forms a multi-bit frame intensity image, where each pixel is composed of an aggregation of jots. Summation is performed in an image signal processing (ISP) block requiring a frame store, while a memory location is required per pixel to sum to the required output frame image bit depth. This has the downside that every doubling of frame image bit depth halves the output frame rate. To address this, real-time on-chip adaptive oversampling is the ideal target of this imaging concept. SPADs were considered by Fossum as the detection device but device scaling issues toward nanometre pitch precluded the technology from further investigation and development. However, SPADs with a distinct voltage spike per photon-induced avalanche makes them an ideal detector to offer a look ahead to the operation and properties of future QIS sensors. This oversampled imaging concept is investigated in this thesis, and is combined with time-domain imaging techniques, to explore gated oversampled imaging, in order to capture high spatial resolution 3D images.

1.2. Research Aims and Background

This research project was conceived to develop architectures for high spatial and temporal resolution CMOS SPAD-based image sensors, in order to create a proof-of-concept gated image sensor for single photon imaging and 3D vision. To achieve this aim, the central proposal was to create a compact, scalable and high-fill-factor SPAD image sensor pixel design. It was envisaged that analogue counting techniques, would provide the most practical approach in facilitating a small SPAD pixel pitch whilst achieving nanosecond temporal accuracy with high dynamic range of photon-counting and single photon sensitivity. Alternative 'QIS' pixel functionality was proposed reducing the full-well down to two discrete logical states (high or low) permitting the capture of a single SPAD event using digital oversampling techniques to form an image. The realisation of SPAD-based image sensor pixels, with low pitch and high fill factor, is the technology jump needed for SPAD technology to be competitive with existing time-domain TOF CIS, and single photon sensitive scientific imaging sensor technology (for example EBCMOS, EMCCD and EBCCD).

To put this research project in context, Prof. Robert Henderson's research group between 2006 and 2011 had two central lines of pixel development for SPAD-based time domain sensing and imaging devices: all-digital image sensor pixels above 44μm pixel pitch, and a plethora of sub-25μm pitch pixel designs for digital silicon photo-multipliers (DSiPM). The group's DSiPM pixels are made compact by including only the SPAD front end and readout addressing circuitry, but do not have an ability to capture, measure and store a photonic event in-pixel. Two examples of such DSiPMs provided targets for small pitch and compact pixel arrangement. Prof. Robert Henderson's own 5μm pitch pixel in ST's 90nm imaging process, with only three NMOS transistors, remains the smallest SPAD pixel pitch in the literature [47]. Dr Eric Webster's 11.6μm pitch DSiPM pixel with near infrared (NIR) sensitive SPAD was a good indicator of a possible small pixel in ST's 130nm process [48]. These prior approaches formed a foundation for, and gave direction to, this research into compact and scalable CMOS SPAD imaging pixels.

At the start of the project, the competing TOF sensors using photo modulated devices (PMDs) were just below QVGA at best (320x200 [49]) and sub-10μm pitch pixels had been demonstrated [50]–[52]. The original sub-20μm pitch target became sub-10μm, in order to be competitive and to demonstrate what was possible with SPADs in ST's imaging 130nm CMOS technology. Previously published SPAD-based TOF image sensors were implemented with considerable in-pixel functionality such as a time to digital converter [40], [53], [54] or multiple ripple counters [55], [56]. However, this inclusion of in-pixel CMOS logic was to the detriment of spatial resolution, silicon area and optical fill factor. At the beginning of this research project, Dr Lucio Pancheri from Fondazione Bruno Kessler (FBK) published the smallest pitch SPAD image sensor pixel, a remarkable 25μm pitch, by employing an analogue counter pixel circuit [57]. These factors reinforced the direction of this research into analogue-based pixel design.

To spark off the research project, three SPAD pixel schematics were created by Prof. Robert Henderson. The first was an analogue counter that formed the basis of this research project. The second and third were time to

amplitude converters (TACs) that were initially developed and trialled and went on to be the foundation of research group colleague Luca Parmesan's doctoral research for FLIM (the initial work into these pixels is covered in Appendix 1 and [4]).

This research project was envisioned to kick start and be a proof-of-concept for several other scientific research projects requiring such high resolution SPAD sensors: namely Luca Parmesan's PhD research into SPAD design and SPAD-based image sensors for FLIM, and Prof. Robert Henderson's ERC Fellowship 'TotalPhoton' grant developing sensors for super-resolution microscopy techniques such as STORM and PALM and FLIM techniques such as FLIM-FRET.

1.3. Contributions to Knowledge

Several contributions to knowledge made in this work are discussed here.

In CMOS image sensor design, it has been challenging to achieve single photon sensitivity due to the difficulty in obtaining a high enough charge to voltage conversion factor (CVF) (e.g. >200μV per photo-generated electron) in a PIN photo-diode and APS circuit with low temporal read noise in the readout. Recently, the first example of a CIS single photon counting pixel with sub 0.3e- read noise has been reported by Prof Eric Fossum's team at Dartmouth University [58]. EMCCD image sensors achieve photon counting through avalanche multiplication in the readout path, but this technique is not compatible with CISs. This thesis illustrates that the combination, of analogue pixel electronics and shared-well substrate isolated SPAD devices, creates such a CMOS photon counting device: a high fill factor SPC image sensor with sub 0.15e- read noise which exhibits photon shot noise limited statistics by exploiting the inherent gain of the SPAD.

The 320x240 (QVGA) time-gated SPAD image sensor, presented in this thesis, is the highest resolution SPAD image sensor yet reported. The 9T NMOS-only pixel design in the image sensor implements a novel method of operating a charge transfer amplifier (CTA) circuit [59]. Using CTA analogue counting pixel operation, the sensor displays global shutter single photon counting images and attains a minimum 0.06e- read noise (taking one SPAD event as the equivalent of one photo-electron). This is an order of magnitude lower input-referred noise than any current scientific or consumer CIS. However, the maximum counting capability or effective 'full well' is limited to under approximately 125 counts where only the first eight to ten captured photons are shot-noise limited due to the accumulation of temporal noise in the counting operation.

The first examples of sub-20μm SPAD-based image sensor pixels at both 9.8μm and 8μm pitch are demonstrated, with 3.1% fill factor in the former and improved to 26.8% in the latter. The latter is the smallest and highest fill factor pixel for a SPAD-based image sensor to date. This attains both the lowest number of transistors for a time-gated SPAD image sensor pixel and the first NMOS-only pixel design without a fixed bias current. Three well sharing techniques are proposed to minimise the physical area occupation of the SPAD

guard ring. Using the most aggressive well sharing technique allowed the pixel pitch reduction to 8μm while increasing the photo-active area by 5.8 times. These reductions in pixel power consumption and in physical area, demonstrate, for the first time, a CMOS SPAD-based image sensor pixel architecture that is scalable to megapixel arrays.

A second pixel operation mode is proposed, functioning as a time-gated photon-triggered single bit dynamic memory. Using the image sensor in this binary image capture mode, reaching 20,000 binary frames per second readout, it offers a look-ahead to the capabilities of future QISs. The bit density (or cumulative binary intensity) against exposure relationship, of the image sensor, is in the shape of the famous Hurter and Driffield densitometry curves of photographic film [60]. This is confirmed through experiment to match the S-shaped 'DLogH' ideal QIS theory curve predicted by Fossum in 2013 [61]. The single bit capture has a bit error rate (BER) of 0.0017 which is an equivalent of 0.168e- read noise.

The QIS mode operates in real-time at 5 to 20kFPS with FPGA-based summation outputting digitally oversampled 8 to 10 bit depth frames at 32 to 8 FPS video rate respectively. In the final chapter, an expansion of the QIS sensor mode is described specifically for STORM and PALM microscopy. Moving-average bit-plane processing is proposed which permits an output image stream at the same kilo-FPS frame rate as the input rate of the sensor. This represents two orders of magnitude improvement in frame rate, versus conventional fixed average bit plane processing, which is critical for the video capture of fast temporal dynamics. Videos are included in the appendices to illustrate this novel image processing technique in comparison to fixed oversampling.

The digital oversampled mode is exploited to capture nanosecond time-gated images, demonstrated through indirect time of flight (ITOF) imaging. In the QIS mode, real-time computation to perform ambient subtraction is applied, and the sensor realises 38mm ranging standard deviation (σ) under a measured 0 to 60cm range.

Alongside the research presented in this thesis, Appendix 2 describes parallel research undertaken into a novel folded flash time to digital converter (TDC) architecture that directly creates a time-domain histogram on-chip (or on field programmable gate array (FPGA)) for single point DSiPM-based TCSPC sensors. The conventional TCSPC process of time-code generation and recording (memory lookup, increment and write) is replaced with a fully parallel system which permits an order of magnitude increase in conversion rate and system throughput (to giga samples per second) over all previously published application specific integrated circuit (ASIC) and FPGA TDC designs. The sensor designed to achieve this has been exploited and published in the application areas of direct TOF ranging [5] and visible light communications (VLC) [9].

1.4. Thesis

This project set out to examine the suggestion that CMOS SPADs could be compactly and efficiently arranged alongside analogue pixel electronics to form scalable pixel building blocks to produce high resolution time-resolved single photon image sensors to be primarily exploited in single photon counting and TOF applications. This section outlines how this achievement is described in this thesis.

This introductory chapter provides an overview of the research background of the directions of single photon image capture and time-domain sensing and imaging.

Chapter two provides an introduction to time resolved imaging with CMOS SPAD detectors. An introduction is provided to direct TOF, or time correlated single photon counting (TCSPC), and indirect TOF, or gated or demodulated imaging. The operation and fundamental parameters of SPADs are put forward. Distinctions are made between CMOS SPAD-based sensors in their different formats: single point sensors, line sensors, imagers and large array sensors.

Chapter three begins with the scaling of different SPAD constructions, and their varied integration requirements are compared. SPAD guard ring minimisation and existing well-sharing techniques are discussed. The choice of SPAD is made to create a high fill factor, but with a low pixel pitch SPAD based imaging pixel. A detailed overview of the literature on SPAD-based image sensor pixels is presented. The merits and demerits of existing pixel designs are weighed to create high spatial resolution single photon image sensors. This chapter describes the motivation in designing an analogue time-gated photon counting pixel.

The fourth chapter describes a test array of CMOS SPAD time gated analogue counter pixels. The initial simulations are presented of the two transistor time gate design and the three transistor analogue counter that are the core of the pixel functionality. The results from the first silicon implementation of these pixels are presented in terms of single photon sensitivity, full well and variability, and are reported in [1].

Chapter five provides a design overview of the SPC image sensor and FPGA control and processing firmware. A revised pixel design is described lowering the pitch to 8μm, removing two transistors and increasing the fill factor to 26.8%. The readout channels and supplementary blocks are detailed. Two readout modes are described – analogue and digital. The correlated double sampling (CDS) analogue readout path to single channel external ADC is used for high sensitivity global shutter image capture. The digital mode reads out single photon, binary bit planes from the image sensor. Reconstructing the image by oversampling in the digital domain on a FPGA is described in the initial publication of this image sensor in [2]. Two types of image capture are performed using these two readout functions: single photon intensity images and indirect TOF images.

The sixth chapter provides the evaluation results of the image sensor for analogue single photon counting and single-bit digitally oversampled single photon counting imaging. These results are published in [2], [17]. The

compact time gate performance is characterised, and ITOF 3D vision is demonstrated using the image sensor with results published in [6].

The final chapter summarises and concludes the thesis and proposes future direction for the work. The success of the image sensor is considered. An examination of the limitations of the pixel, the system and the technology is undertaken proposing a prospective future direction of the research.

2. Time-Resolved Imaging with CMOS SPADs

2.1. Introduction to Time Resolved Imaging

The term time-resolved imaging encompasses both wide-field TOF imaging and scientific TCSPC (predominantly single point scanned systems). TCSPC relies on the same concepts as direct TOF and for the purposes of this introduction they are considered together.The reader is directed towards an excellent book on TOF range imaging and 3D vision 'TOF Range Cameras' [62] which also covers TOF illumination techniques. Moreover, Appendix 5 presents a literature review on 3D vision techniques and Appendix 6 describes further background in time-resolved imaging including an overview of TOF illumination. Figure 2.1 comprises a taxonomy diagram of TOF imaging techniques. The highlighted route in dark blue from TOF to analogue counters and digital memory shows the central area of research in this thesis. The light blue highlighted routes show areas that were investigated but are outside the core work in this thesis and are described in Appendices 1 and 2, and in co-authored publications and patents listed at the beginning of this document.

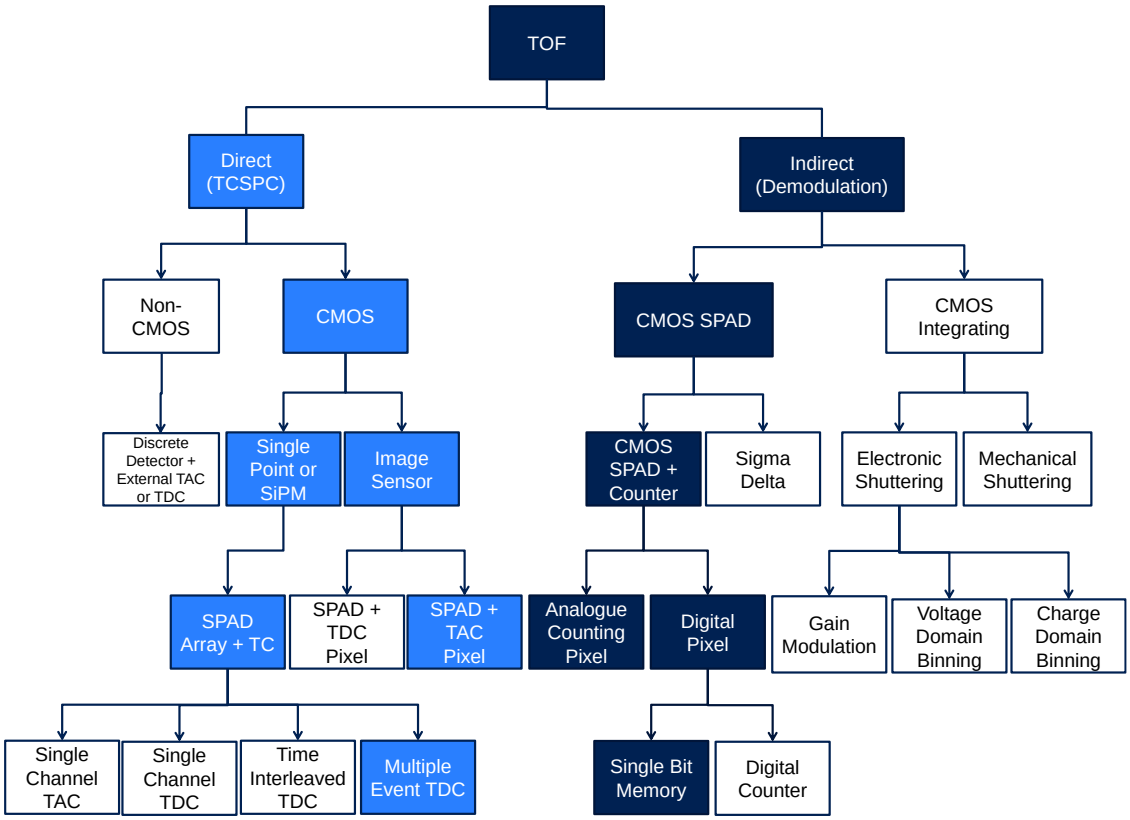


Figure 2.1. Taxonomy of TOF detection techniques. The dark blue highlighted path describes the research presented in this thesis. The light blue highlighted areas were investigated as part of this research but do not form the core of this thesis.

2.1.1. Direct Time of Flight

The first experiment to determine the TOF of a light pulse was performed in order to calculate the speed of light. This experiment predates this thesis work by more than 125 years; physicists Fizeau and Foucault in 1849, placed a mirror and lens at a distance of a mile, then set up a rotating cog wheel with narrow slits at their observer position. They shone a beam of focussed sun light through the cogwheel slits, via focussing optics, to the mirror and back. They tuned the cogwheel rotational speed until the light pulse shining out through one cogwheel slit returned and passed through the next. Using the following equation, they determined the speed of light to be $3.15 \times 10^{8} ms^{-1}$ which was remarkably accurate:

$$c = \frac{2 \, . \, D \, . \, v}{d}$$

(Eq.2.1)

where D = distance to mirror, v = cogwheel rotational velocity, d = slit separation. In 1922, Marconi first reported his discovery of detection of objects using radio waves [63]. This led to the WWII innovation of radio detection and ranging (RADAR), measuring the TOF of a short radio pulse from transmission, reflection off a metal object and back to a receiver. This direct measurement of the TOF of a short pulse forms the basis of optical or light detection and ranging systems (LIDAR). A laser pulse with ns to ps duration, is transmitted, reflected and detected by a fast photodetector. The electrical pulse from the photodetector is input into an electronic 'stopwatch' circuit – started co-incident with the laser pulse and stopped with the received signal (or vice versa started with the received signal and stopped with the next laser signal). This direct TOF (DTOF) measurement circuit creates a timestamp for every detected pulse as shown in Figure 2.2.

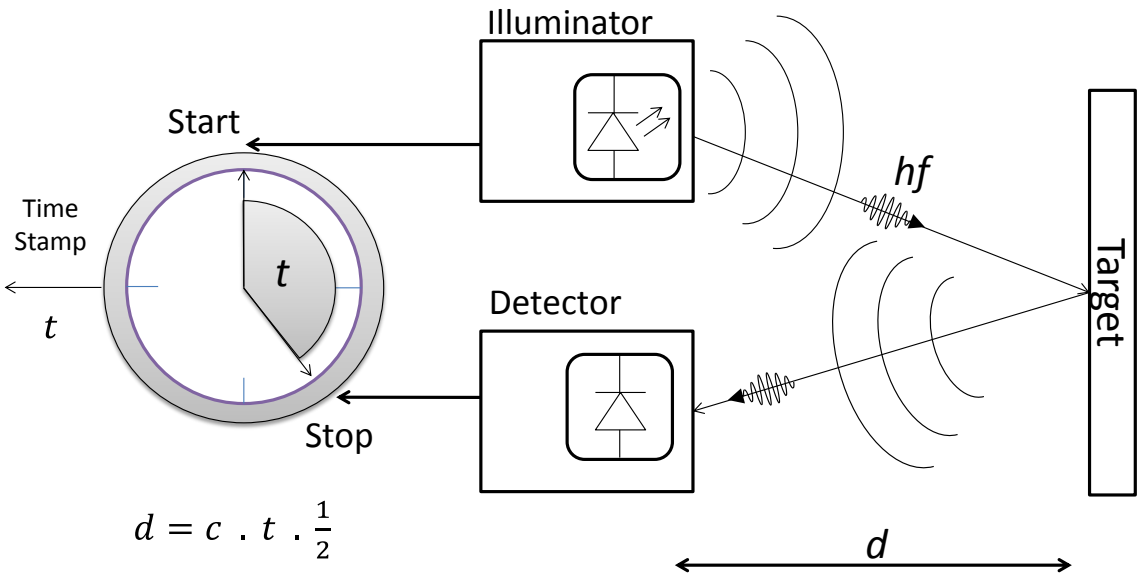


Figure 2.2. Direct time of flight measurement showing forward mode or start-stop timing.

These timestamps are subsequently processed either by averaging [64] or by creating a time-domain histogram of the optical waveform [5], [30]. The histogram creation is conventionally performed by a memory lookup (using the timestamp as the memory address) and increment of the memory location representing one bin of the temporal histogram [30]. When combined with a single photon detector, the latter approach is referred to as time correlated single photon counting (TCSPC). Figure 2.3 illustrates the TCSPC concept for a single detector starting the time conversion circuit and the succeeding laser synchronisation pulse stopping it. This is referred to as reverse mode or stop-start timing as a time recorded is the remainder of the period between the 'sync' pulses, in contrast to the forward mode timing shown in the previous figure. For detailed information on TCSPC systems and methods, the reader is directed to Becker's excellent textbook on the subject [30].

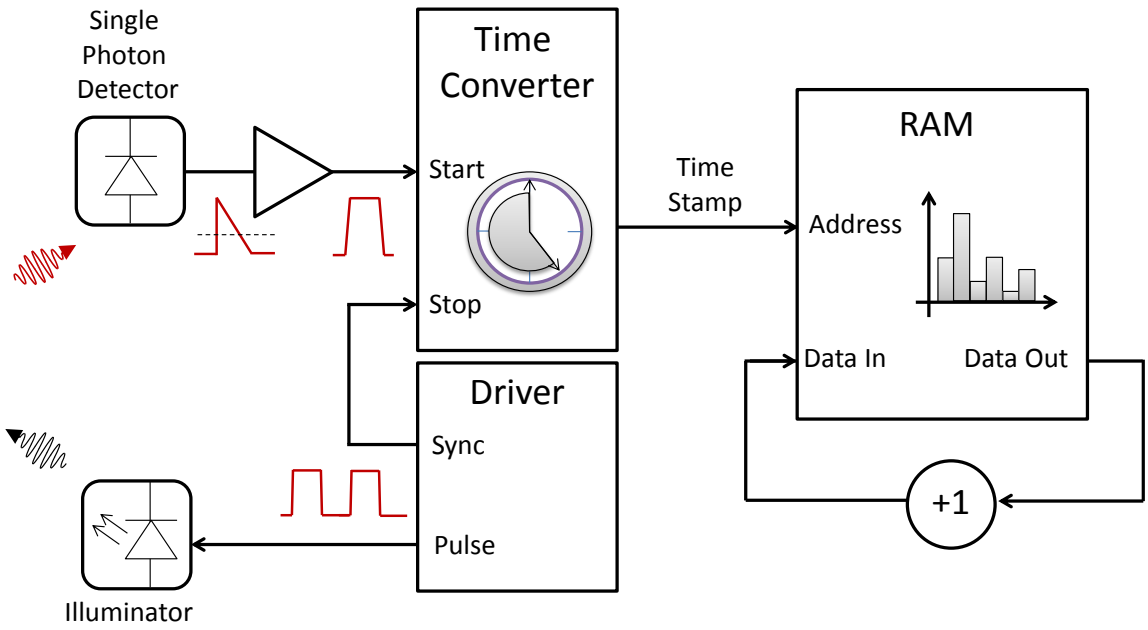


Figure 2.3. Time correlated single photon counting block diagram showing reverse mode timing.

To create a DTOF image sensor using such TCSPC architectures is a technical challenge. It is a highly memory intensive procedure; each pixel creating a time-stamp for every photon arriving, then processing these into a per-pixel histogram. Where the high data rate and external memory are acceptable for low resolution scientific imagers, the data transmission rate and memory requirements indicate that TCSPC image sensors are not scalable for megapixel TOF imaging arrays in this current form.

2.1.2. Indirect Time of Flight

In contrast to DTOF or TCSPC, the TOF may be resolved by measuring the relative phase shift ($\Delta\varphi$) of a detected signal with respect to an illumination signal. This phase measurement does not directly determine the absolute TOF, but instead infers the temporal offset or delay of the received signal through a calculation using the frequency or time domain phase information, hence the term indirect TOF (ITOF). The notable advantage of ITOF over DTOF is that the phase measurement may be performed entirely in-pixel whereas DTOF systems require a TCSPC histogram to be created per pixel which, as previously mentioned, is a huge overhead in terms of data management and silicon area for per-pixel storage. Conventional ITOF systems have the significant advantage that they read out like a conventional image sensor, through an ISP pipeline, with no external per-pixel storage and frame store.

Figure 2.4 illustrates the principle of ITOF system where a single frequency (f) amplitude modulated continuous wave (AMCW) illuminating light is directed towards an object and the frequency domain phase shift ($\Delta\varphi$ equivalent to a time-domain time delay 't') of the returned signal is measured and the distance to the object is calculated using the following expression:

$$D = \frac{c \,.\, \Delta\varphi}{4 \,.\, \pi \,.\, f}$$

(Eq 2.2)

A pulsed illuminator system is similar, where the measurement occurs in the time domain. Using two time-gated windows a time-delay can be inferred by the relative intensities or values of the two recorded signals. Additional time-gated windows without pulsed illumination can be used to determine the signal offset caused by background light in order to subtract it. Figure 2.5 illustrates two bin time-gated window ITOF measurement. For four (or more) bin pulsed ITOF the reader is directed to [36], [62], [65], [66] as examples.

The limitation of those ITOF systems using a single frequency of illumination and sampling with two or four bin samples is that only a single phase or time delay measurement is captured. If two spatially separated objects are detected by a single pixel then the weighted complex sum between the two objects is calculated as the distance [62]. For example, if an AMCW ITOF camera images an object through a glass window then it will capture a distorted average of the glass and object whereas a DTOF system (with appropriate processing) is capable of resolving both glass and object independently. Gobadz et al. gives detailed treatment of this distortion due to multiple returns in [62].

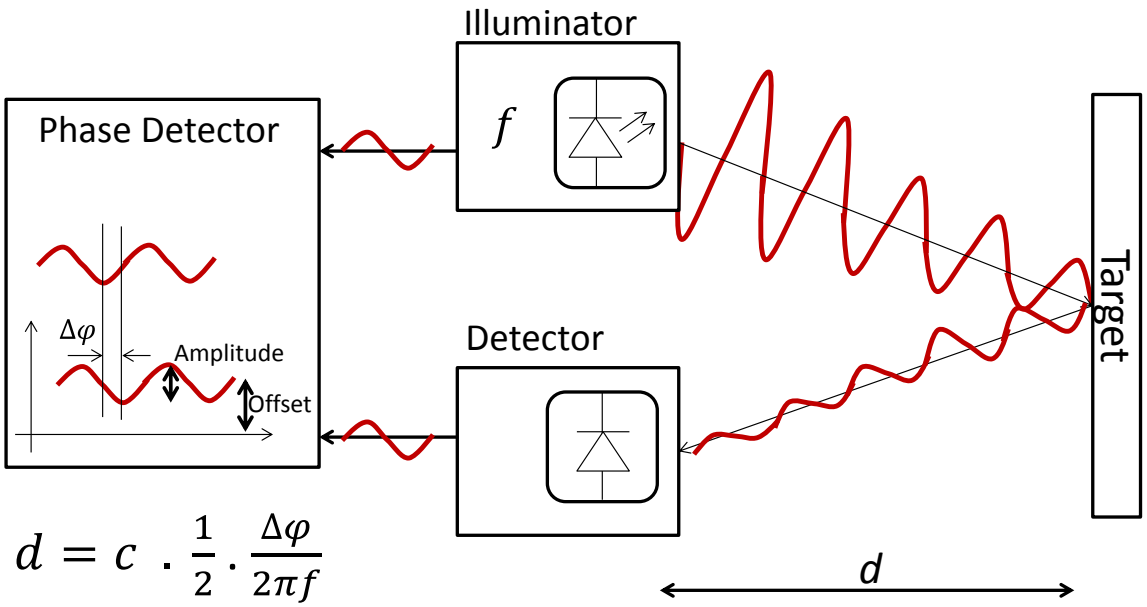


Figure 2.4. AMCW ITOF showing phase, amplitude and offset of the detected signal.

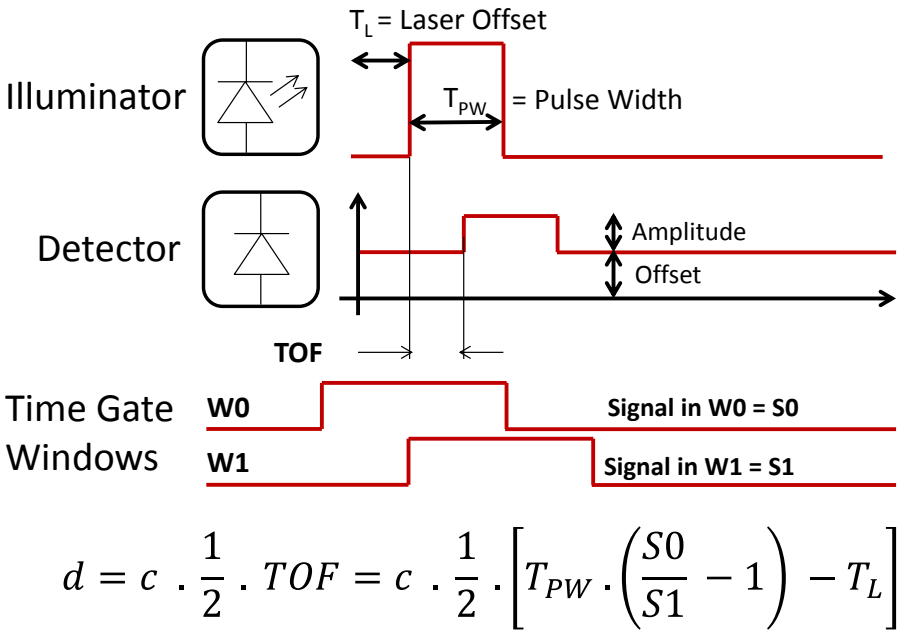


Figure 2.5. Pulsed illumination ITOF captured with two time-gated windows.

2.2. The CMOS Single Photon Avalanche Diode

This section will briefly introduce and discuss the operation of the CMOS SPAD. The reader is directed to two thorough reviews on SPAD device physics, integration in CMOS, operation and history in [67] and [68].

2.2.1. CMOS SPAD Introduction

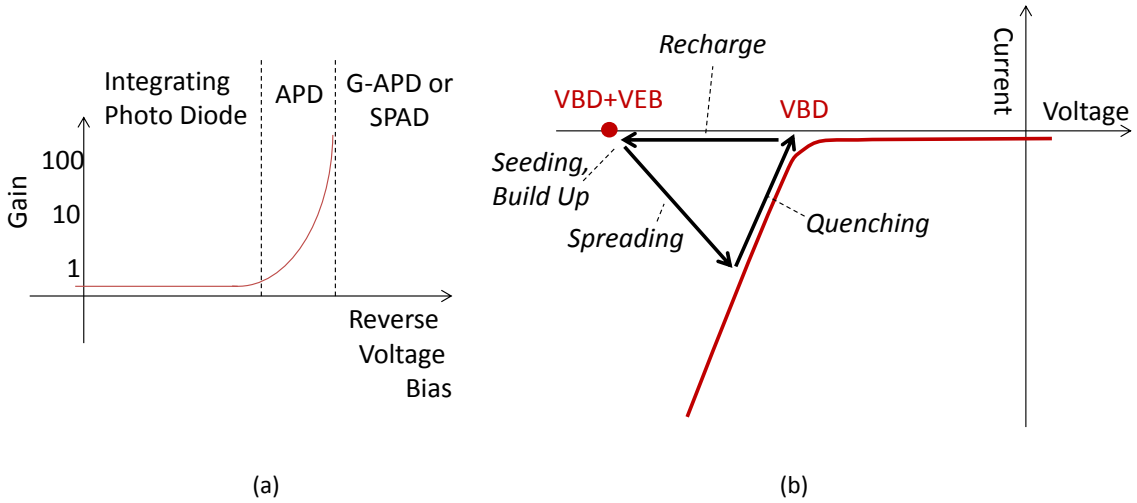


Figure 2.6 (a) Gain to reverse bias plot describing the regions of photo-diode operation. (b) The I-V plot showing the five stages of SPAD operation.

The SPAD device is a subset or class of avalanche photo diode (APD) operating in 'Geiger' mode (G-APD). The three regions of photo-diode operation are illustrated in the 'gain to reverse bias' graph in Figure 2.6(a): integration, avalanche and 'Geiger' single photon avalanche mode. The SPAD is a depleted PN junction biased and operated above its reverse breakdown voltage (V_{BD}) with a high electric field across it. Upon impact ionisation of a photon into an e-h pair, the charge (whether e- or h) entering the depletion region triggers a current avalanche. The probability of this avalanche occurring is determined by the ionisation rate controlled by the excess bias (V_{EB}) above V_{BD} applied to the junction which increases the electric field and increases the ionisation rate. The point of operation is quasi-stable above breakdown (until an avalanche is triggered), and is shown by the red dot on the I-V plot in Figure 2.6(b). The triggered avalanche process has four stages before the final fifth recharge stage:

- Seeding: the introduction of a free charge, from the ionisation of a photon, into the diode depletion region. The probability of the ionisation of free charge carriers in the depletion region exceeding unity is the necessary condition for the avalanche process starting.

- Build-up: a rapid process whereby the local current density increases and the local voltage around the seeded diode region (internal to the diode and not observable externally) begins to fall from the excess bias ($V_{EB} + V_{BD}$) to the breakdown voltage.
- Spreading: the avalanche current spreads through a diffusion process and multiplies across the entire diode. This process is externally observable.
- Quenching: the avalanche is halted due to the rise of the potential of the moving node (either anode or cathode) as the voltage across the diode has decreased to the breakdown voltage. The quenching process is intrinsic to the diode [69], but actively assisted (transistor based) quenching has also been investigated for large area diodes [70]. At this stage, the SPAD is no longer photo-sensitive.

After quenching, a passive or active circuit begins the process of recharging the diode. As shown in Figure 2.7(a), this is accomplished by a resistance (or transistor) restoring the excess bias and electric field (and consequently photo-sensitivity) across the diode, ready to begin the avalanche process again. The SPAD creates a distinct voltage spike for each avalanche event, which is buffered or amplified by a MOS circuit. SPADs whose avalanche causes the moving node (anode or cathode) potential to increase from low to high and then recharge back to a low potential are referred to as positive drive devices. Conversely, those devices which have a negative-going drop in potential of the moving node are negative drive SPADs. The leading edge of the voltage spike, shown in red in Figure 2.7(b), signals the arrival of a single photon with a temporal precision of 10's to 100's of picoseconds [68]. Like Geiger radiation detectors determine the intensity of radiation, the intensity of the incoming light can be measured by counting the output pulses (hence the alternative term GAPD). The duration of the avalanche and recharge is known as SPAD dead time, as the detector is dead or unresponsive to incoming photons. It is controllable by the recharge resistance and is in the order of nanoseconds. Minimisation of SPAD dead time is a key design goal to maximise count rate.

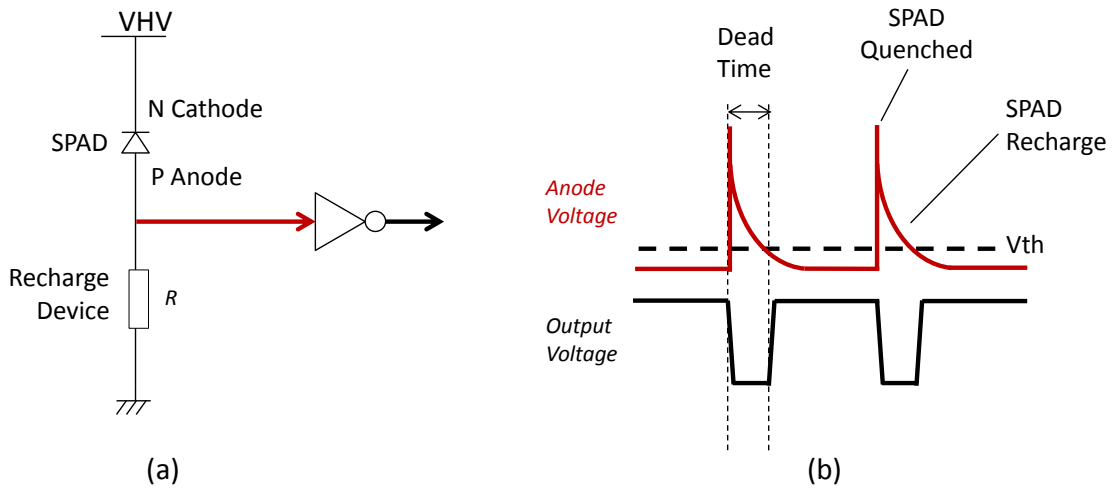


Figure 2.7 (a) SPAD passive recharge circuit with CMOS output inverter. (b) Example of anode and output waveforms of a positive drive SPAD.

Vertical and 3D cross sections of a CMOS SPAD diode structure are illustrated in Figure 2.8. All diode structures can be made to avalanche – the challenge lies in designing a non-destructive avalanche diode with a wide and uniform electric field creating an even photo-sensitive area. As illustrated in the dotted red regions, the SPAD structure is intended to have a uniform laterally planar multiplication and breakdown region. The surrounding cathode ring is designed to create a guard ring structure ensuring that breakdown does not occur at the vertical edges of the anode, thus preventing a sensitive annulus region.

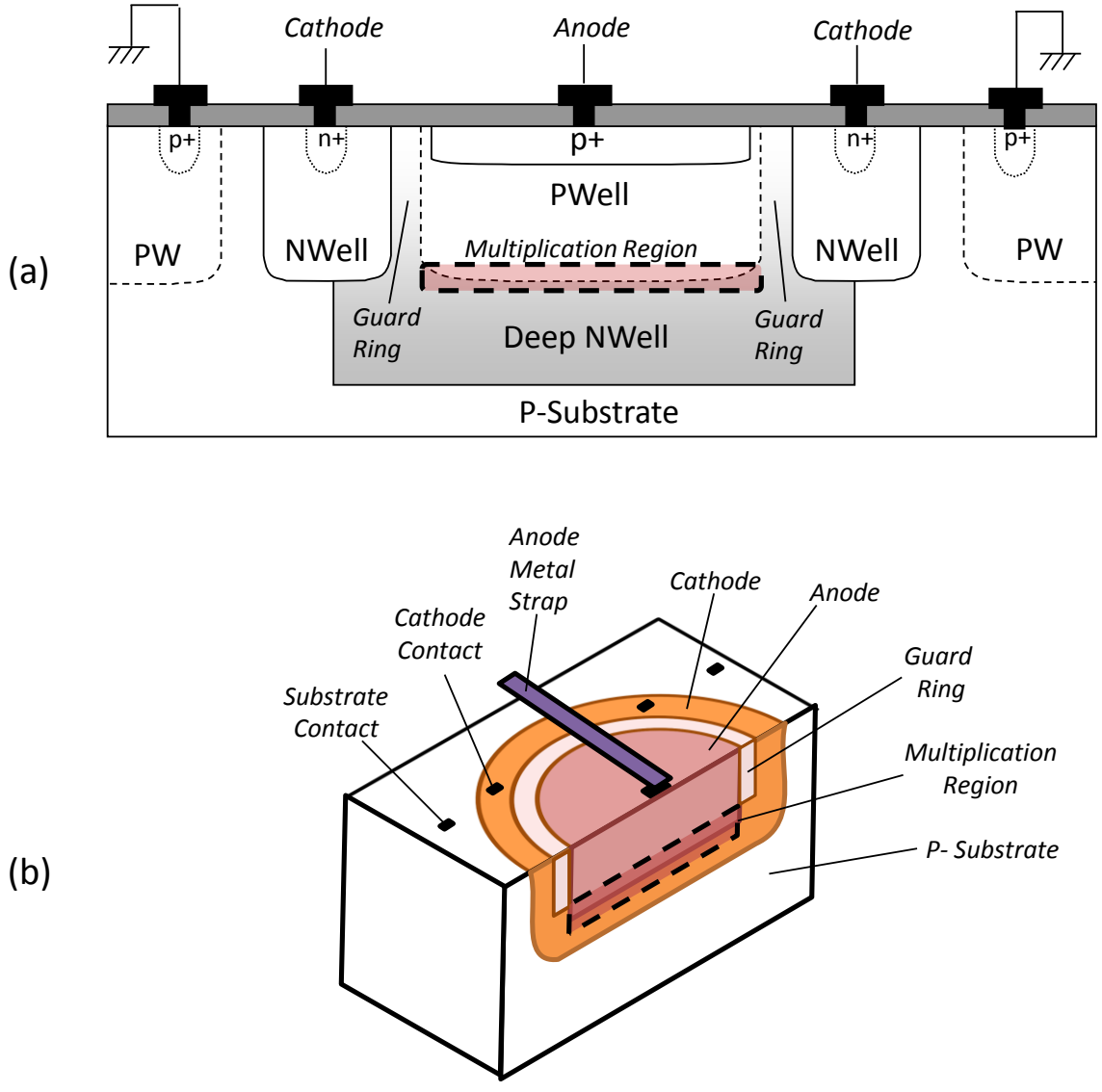


Figure 2.8 (a) Vertical cross section of a CMOS SPAD structure. (b) Simplified 3D SPAD cross section.

2.2.2. CMOS SPAD Performance Parameters

There are three key performance parameters of CMOS SPADs: sensitivity, noise and temporal resolution.

Sensitivity is measured by photon detection probability (PDP): a combination of the intrinsic quantum efficiency (IQE) of the diode structure and the avalanche probability determined by the excess bias above breakdown. Photon detection efficiency (PDE) is defined by the product of PDP and FF, and is the SPAD equivalent of extrinsic QE (EQE). The SPAD recharge duration or dead time, affects the maximum number of photons that can be detected within an exposure period, and therefore impacts sensitivity.

Noise is caused by thermally generated carriers, or defectivity in the form of charge traps, or band to band tunnelling from within or through the guard ring. This dark noise generates SPAD avalanches unrelated to the photon flux and is measured as the dark count rate (DCR) in counts per second. DCR has a dependence on the temperature and excess bias. Traps can store a charge from a prior avalanche and release it after a period of time creating a secondary correlated avalanche referred to as an after pulse, another SPAD noise source.

Temporal resolution or jitter is a measure of the statistical variation in the onset and detection of the SPAD avalanche from the impact ionisation of a photon. It is defined either as the standard deviation (or RMS) value, or as the full-width half-maximum (FWHM), of the recorded pulses given a repetitive optical impulse stimulus (from a picosecond or femtosecond laser).

2.3. CMOS SPAD-based Sensors

Up to 2015 there have been 13 years of research and development in CMOS-integrated SPADs by a small number of academic groups, and more recently companies (both with fabrication plants and 'fabless'). Alexis Rochas, then at EPFL, reported in 2003 the first SPAD integrated in CMOS and subsequently the first fully integrated CMOS SPAD array [71], [72]. Since then many SPAD-based sensors have been designed and manufactured in CMOS for a range of time resolved and single photon counting applications. Unlike CISs with the APS-based readout [73], there is not currently a 'one size fits all' solution for any one particular SPAD-based pixel architecture or SPAD sensor variety. This has led to many SPAD sensor architectures and a plethora of pixel designs. They can be broadly classed into four general categories and Table 2.1 provides an illustrative view of some of the distinguishing parameters between these four architectures:

- Single point or digital silicon photo-multipliers (DSiPM).
- Line sensor.
- Image sensor.
- Large array sensors specifically for PET.

Single point sensors are created from arrays of SPADs logically combined together to create a large area single point detector. They provide temporal resolution in the 50-100 picosecond regime, ideally with high output data rate or throughput. DSiPMs have been described which either provide a stream of time-stamps for off-chip histogram generation [74] or alternatively perform a histogram on-chip [5]. All DSiPM pixels and some line sensor pixels achieve high fill factor arrays through placement only of the quench and recharge, and addressing circuitry, in-pixel.

Line sensors have been published in the literature with up to 1024 SPADs in the X dimension [75], but are limited in scaling in the Y-dimension as the per-pixel photon counting or time circuitry is placed column parallel below the imaging array. Where 50ps temporal resolution has been described [76], the sensor throughput becomes limited as individual time-stamps require to be streamed out. Streaming this data off-chip creates a bottleneck and leading to pile-up distortion [77]. Processing and compressing this data on-chip addresses this limitation [78]. Furthermore, chip stacking technology will address this bandwidth limitation [79].

SPAD-based imagers have the additional in-pixel capability to capture, measure and store a photonic event. There are two distinct classes of SPAD image sensors. Those with high temporal resolution are limited in array size, whereas those with high spatial resolution are limited to nanosecond time gating. High temporal resolution (approximately 50ps) imaging sensors have been described but are limited to low array sizes as again, data output bandwidth is limited. Image sensors which attempt to attain both high temporal and spatial resolution are severely limited in output bandwidth even with large numbers of parallel output channels [54], [56].

Large array sensors have been described for the photon-starved application of TOF positron emission tomography (PET) [80]–[83] and endoscopic TOF PET [84]. Large array PET IC's are designed to be abutted on four sides to create a PET ring detector [41]. In each sensor, arrays are constructed of DSiPM macro-pixels, each consisting of SPAD array, photon counters and high temporal resolution TDC [85]. The sensor throughput is intentionally limited due to the full system data processing constraints with many detectors reporting simultaneous events.

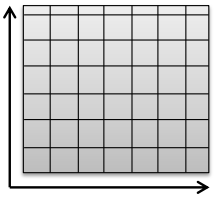
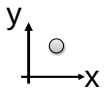
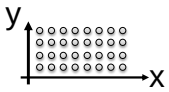
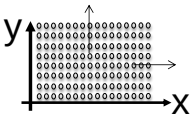
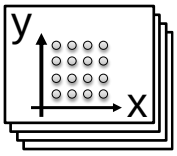
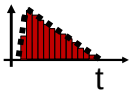
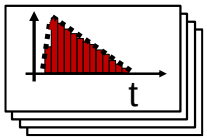
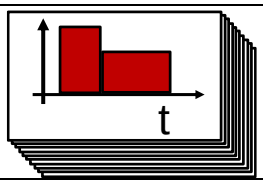
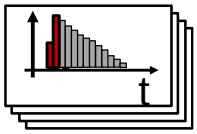
Parameter	Single Point (DSiPM)	Line Sensor	Image Sensor	PET Sensor
Chip Diagram (Dark grey denotes light sensitive area)				
Spatial Resolution Diagram				
Spatial Resolution	Single Point	Many X, Few Y	Many X & Y	Low spatial resolution per sensor with arrays of DSiPM macro-pixels.
Max Array Size	32x32 [5], [74]	1024x8 [75]	320x240 [2]	2x2 [85] 16x16 (SPADNET2 Unpublished)
Spatial Scalability Limit	Limited Scaling in both X & Y	Limited scaling in Y	Data rate limits scaling	Requires 4-side buttable IC's for tiling to create large area arrays.
Temporal Resolution Diagram				
Temporal Resolution	Single channel with ps resolution	Multiple parallel channels with ps resolution.	ns for high spatial resolution ps for low spatial resolution	Multiple parallel channels with ps resolution.
Sensor Throughput	High	Limited by output data bandwidth	Limited by output data bandwidth. Requires in-pixel integration for high throughput.	Bandwidth limited: the first photons in the photo-emission burst or flash encoded.

Table 2.1. Four general categories of SPAD-based time resolved single photon counting sensors.

2.4. Chapter Summary

Time resolved imaging has been introduced in this chapter with a taxonomy of TOF and time resolved detection techniques focussing on CMOS SPAD based devices. Direct TOF and TCSPC methods were presented and the difference between forward and reverse start-stop timing is discussed. Recording time-stamps for every incoming photon is a memory intensive procedure and is not scalable to a wide-field image sensor. Gated and indirect TOF imaging, with in-pixel functionality and storage, permits wide-field image capture of time-resolved optical events at the expense of detailed time-domain information.

A preliminary explanation was given of the operation and fundamentals of CMOS SPADs. The sensors based on this technology may be generally categorised into single point sensors, line sensors, imagers and tile-able large array sensors specifically for PET. The two primary limitations to the scalability of each of these architectures, are data bandwidth (both data processing and off chip transmission) and pixel construction. The next chapter discusses both the SPAD and pixel electronics in the search for a scalable pixel design to attain a high spatial resolution SPAD-based image sensor.

3. Single Photon Avalanche Diode Pixel Design

3.1. Introduction

The first part of this chapter presents an overview of the scaling of the many different and possible diodes permissible in the ST Imaging CMOS 130nm process. A number of factors are considered: SPAD to circuit coupling, passive and active quench and recharge, and the geometry of pixels and diodes. Different SPAD constructions are summarised and original concepts in diode well sharing are presented.

In the second part of this chapter recently published CMOS SPAD sensors are discussed in terms of pixel design. Pixel structures are categorised and compared to realise the efficiency and effectiveness of the respective designs in an image sensor context. The review is presented in terms of pixel scalability, in-pixel functionality and the pixel pitch versus fill-factor. Where limits exist in particular pixel architectures in a monolithic process, the alternative chip stacking pixel architectures are considered. A conclusion is made to select the optimum pixel for a time gated single photon counting SPAD-based image sensor.

3.2. SPAD Choice and Implementation

This section details the scaling of different SPADs, and the accompanying quench and recharge circuits, down to minimum pixel pitches. In the first part of this section, the circuits required to implement a SPAD are considered in terms of applying them in a scalable pixel format to implement a high resolution image sensor. In the second part, the many factors of implementing a SPAD are considered: scaling the diode and guard ring down to small pitches, amenability to well-sharing, and the effective PDE. The requirements and factors impacting the choice of SPAD best suited to being implemented in an image sensor context are reviewed and a choice of SPAD is made. These factors are:

- Quench and Recharge Circuits
- Enabling and Disabling a SPAD
- Directly Connected or Capacitively Coupled SPADs

Circuit Implementations

- Guard Rings and Well Sharing
- Avalanche Junction Design
- Breakdown Voltage
- PDE and Jitter
- SPAD Comparison
- SPAD Scaling and Geometry

Diode Implementations

3.2.1. Quench and Recharge Circuits

The quench and recharge circuit can be as simple as a single passive resistor or transistor (termed passive quench) or it can comprise a monostable circuit with different paths for quench, recharge and precharge of the SPAD moving node. There are different combinations of quench and recharge circuits namely:

- Active quench, active recharge (AQAR)
- Passive quench, active recharge (PQAR)
- Active quench, passive recharge (AQPR)
- Passive quench, passive recharge (PQPR)

The applicability of the first three categories with active quench (AQ) or active recharge (AR) circuitry in monolithic SPAD image sensors is considered first in this section, then the single transistor passive quench (PQ), passive recharge (PR) is reviewed.

3.2.1.1. Active Quench and Recharge Circuitry

Figure 3.1 illustrates the three variants of AQ and AR circuits and the voltage waveforms of the SPAD moving node in the three cases for a positive drive SPAD. AQ circuits operate by detecting the onset of the avalanche and force the SPAD moving node to the excess bias supply ('VAQ' in the figure). Whereas, AR circuits detect both that the avalanche has occurred and that the SPAD node has reached the excess bias supply. It operates by pulling the SPAD moving node back to re-activate or re-arm the detector ready for the next photon. Active recharge circuitry is applied in situations where a SPAD needs to be recharged faster than the passive quench 'RC' time constant allows. This is suitable for large diodes, as seen in the extensive work by Cova et al. at Polytecnico di Milano [70].

AQ circuits reduce after-pulsing by holding the SPAD moving node in a disabled state for a known duration after an avalanche [86]–[88]. Moreover, during the recharge of a positive drive SPAD, the bias of the moving node ramps back down from the excess bias supply ('VAQ' in Figure 3.1) to ground. In this period the sensitivity of the SPAD increases from zero to the level set by the SPAD excess bias. If a photon, dark event or after-pulse triggers the SPAD during this time, causing a secondary avalanche, but before the SPAD moving node voltage has crossed the threshold of the front end buffer, then the output of the SPAD buffer appears to broaden in time – a merger of the two avalanches. This manifests as a paralysis of the SPAD in terms of count rate. An active quench minimises the recharge time, reactivating the SPAD in the shortest possible time, effectively mitigating the paralysis seen in passively recharged devices at high light levels [89].

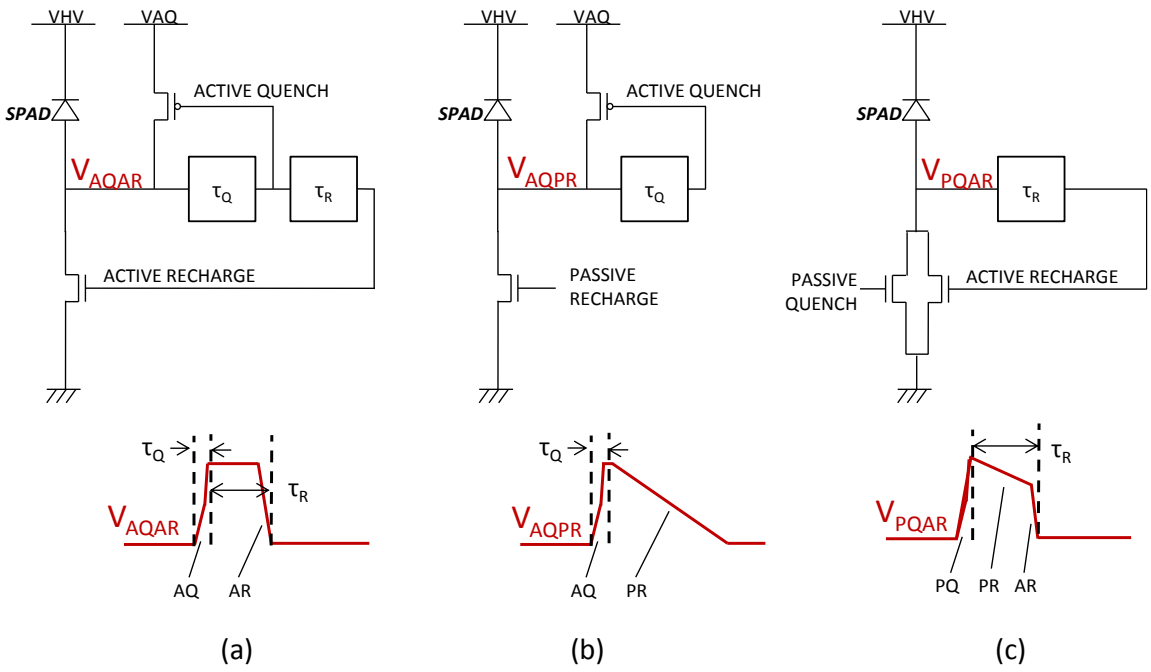


Figure 3.1. The three variants of active quench and recharge circuitry and moving node voltage waveforms. (a) Active Quench, Active Recharge [89] **(b) Active Quench, Passive Recharge** [90] **(c) Passive Quench, Active Recharge** [87]**.**

The area of such AQ or AR circuitry places physical limits on lowering the SPAD pixel pitch by introducing a greater number of transistors than a single transistor PQPR. Leveraging stacking technology would permit the active quench to be placed under the SPAD, eradicating the area constraint.

The stability of multiple active quench circuits operating simultaneously in close-proximity on the same supply has yet to be studied. It is apparent that the monostable circuit, with three or more digital gates, is a fast switching device which will induce high amounts of temporal switching noise onto the supply and into the substrate beside the pixel. Without low impendence power straps and appropriate decoupling, this switching will create a localised power supply drop prolonging the propagation delay of each monostable gate and the delay of surrounding active quench circuits. In applications whose primary goal is sustaining high count rates that can afford the increased power consumption (from AQAR logic switching) for a small-scale array of pixels, e.g. optical communications [91], an active quench is appropriate. However, the high power consumption is clearly an untenable situation for certain applications that demand low power consumption, for example consumer electronics.

3.2.1.2. Passive Quench Passive Recharge

A PQPR device is realised in CMOS by a poly-silicon resistor [67], an NMOS [92] or a PMOS transistor [91] as shown in Figure 3.2 below. Thick oxide transistors are used when the excess bias of the SPAD exceeds the

voltage rating of the thinner oxide transistors in a process. PMOS transistor quenches are used for those negative drive SPAD devices requiring a cathode connection pulled up to an excess bias supply ('V_{EB}') as shown in Figure 3.2(c). Conversely, as shown in Figure 3.2(b) an NMOS quench is used to connect positive drive SPAD devices with anode connections pulled down to ground. These three methods of SPAD connection are fundamentally different in terms of well sharing, and therefore the creation of a scalable compact SPAD pixel is dependent on the choice of SPAD connection. Figure 3.2 (a) and (c) both rely on using the DNW as the moving node, so cannot be integrated in a well sharing approach and are not isolated from the substrate and will incur an increased jitter tail from deep diffusion carriers. The reader is directed to Cova's work on SPAD quench and recharge design in [93] which highlights a rule of thumb of 50kΩ per 1V of excess bias applied to the SPAD which must be considered when choosing a PQPR device.

A hybrid approach is seen in [94] using a poly-resistive passive quench in line with minimum sized PMOS and NMOS switches to allow quenching and disabling pull-down respectively. This combined series-connected method may be revisited in the future as SPAD pixels shrink and designers are forced to use smaller geometry transistors, and yet require high-valued resistance. Although poly-resistors are physically larger than MOSFET devices with comparable resistance, they can be physically placed over the SPAD guard ring (or indeed the SPAD active area) to minimise the area penalty [67]. However, a transistor based PQPR is preferred over a resistor, as the resistance of the transistor can be tuned by the gate voltage to control and minimise the SPAD dead time.

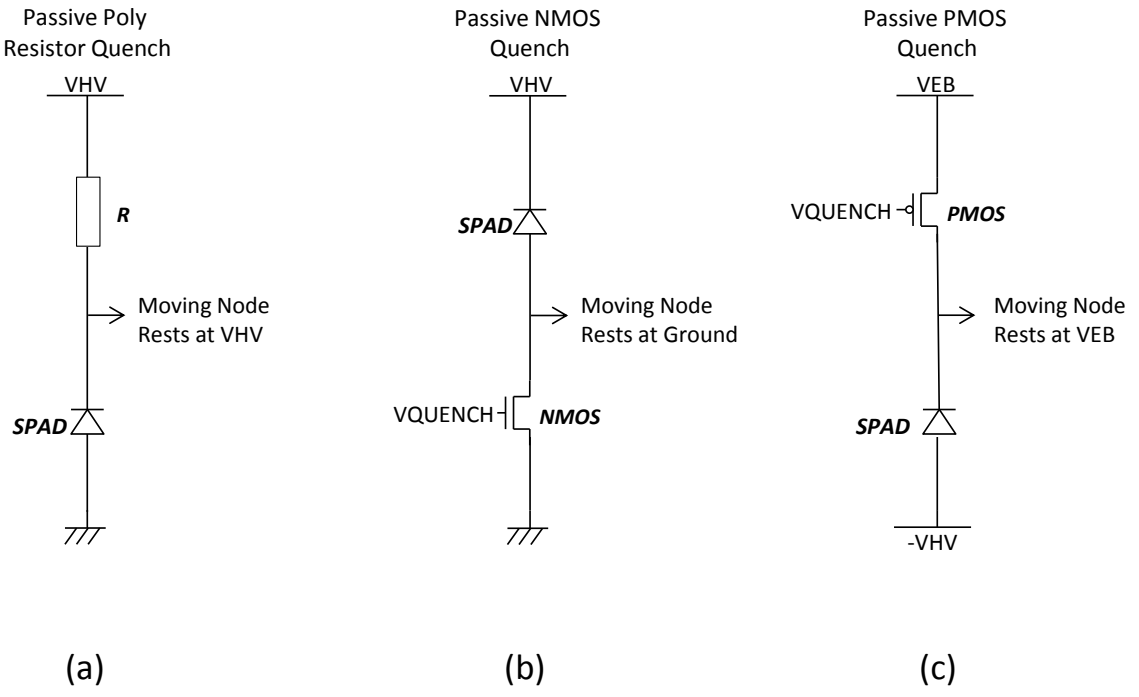


Figure 3.2. Three passive quench, passive recharge implementations: (a) poly silicon resistor (b) NMOS transistor (c) PMOS transistor.

Furthermore, the choice of an NMOS over a PMOS device is preferred for two reasons: the inclusion of an NW increases the circuit area (due to NW spacing requirements from NMOS transistor active regions) and the hot NW spacing rules pushes the PMOS NW away from the SPAD NW making the NMOS-based circuit physically smaller to implement. Furthermore, it integrates with the rest of the NMOS pixel electronics. In summary, a single NMOS transistor PQPR provides the preferred approach for a scalable image sensor pixel over the use of PMOS (with its inevitable increase in pixel pitch and loss of fill factor), or the use of a resistor with no dead time control.

3.2.2. Enabling and Disabling a SPAD

In applications where power consumption is a priority, high DCR SPADs must be switched off and disabled. For the majority of operations, a high DCR SPAD will consume power and contribute noise to the sensor system. Furthermore, high DCR SPADs may increase the DCR of neighbouring SPADs by optical crosstalk [95]. Figure 3.3 illustrates two example circuits to enable and disable a SPAD. Such an enable or disable function is operated by activating or de-activating the PQPR transistor. Moreover, in some examples a switch is used to pull the anode to a supply greater than the excess bias, lowering the potential across the SPAD to below the breakdown voltage.

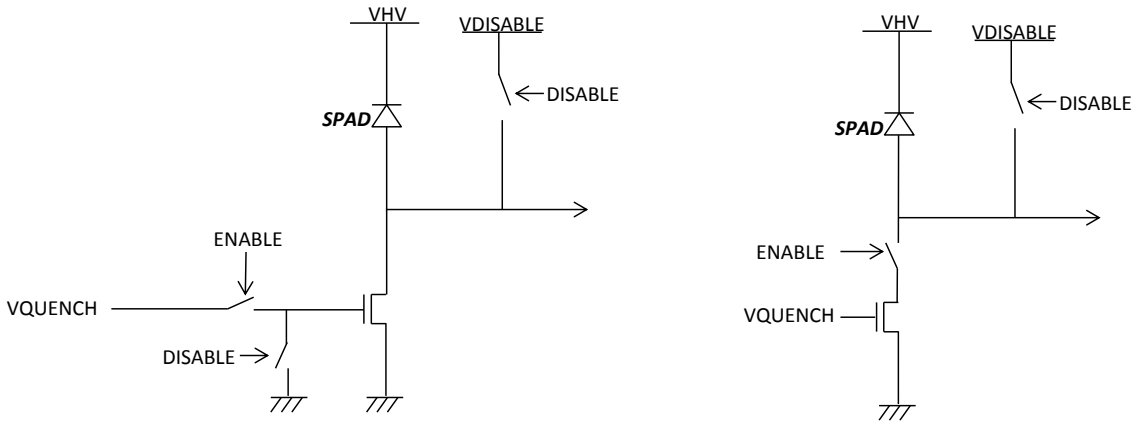


Figure 3.3. Two examples of SPAD enable and disable circuit diagrams showing generalised switches and a passive quench and recharge NMOS transistors where 'VDISABLE' is greater than the SPAD excess bias.

Time-gating the SPAD by this method of excess bias modulation is possible but has difficulties in implementation. The advantage of this time-gating method is the SPAD is insensitive to incoming photons outside the time gate and therefore saves power due to suppression of SPAD activity outside of time-windows of interest. However, this method will incur numerous issues in implementation, first it requires modulating (charging and discharging) the capacitance of each of the diodes in the array which limits scalability of this

technique due to power. Second, the diode must be held in photo-diode mode during its disable period then moved through the APD region before entering the Geiger region – any diode incurring an avalanche during the transition through APD mode will alter and create temporal noise in the time-gate window.

There are three methods to distribute the signals for controlling the activation of SPAD pixels, each with their own impact in pixel design (due to the number of in-pixel transistors) and subsequent pixel area and routing:

- Individual routing from the edge of an array
- XY grid with in-pixel combinational logic.
- In-pixel memory

The impact, of each of these three signal distributions, is considered. First, an enable and disable signal may be individually routed to each pixel. This of course is only applicable in a small array as the challenge of routing many signals from the edge of an array becomes unmanageable beyond eight to sixteen SPADs in a single row. Only the enable signal needs to be routed if the disable signal can be locally generated in-pixel using a CMOS inverter (at the cost of in-pixel transistors and area). This control method is scalable in one axis in a line sensor format and is seen in [35].

Secondly, a grid structure of XY addressing is implemented in [96]. As shown in the example diagram in Figure 3.4(a), such a system operates by asserting a static logic high on a row line and a column line to disable a pixel. Simple CMOS decoding logic is required in the pixel where each disable device requires an 'AND' logic function and the enable device requires a 'NAND' function. This scheme functions only for small arrays where a few SPADs are expected to have high DCR. The scalability limit of this XY scheme is highlighted in the example in Figure 3.4(b) where two functioning pixels (highlighted in dark grey) are incorrectly disabled to ensure the two high DCR SPAD pixels are deactivated. This problem will be significantly compounded with larger arrays.

In single point DSiPM devices, it is conceivable that only a small sub-array of pixels is ever activated out of a large photo-sensitive area (the large area may feasibly only be required for ease of optical alignment and calibration). For this application, the XY grid structure can be extended. The in-pixel decoding logic could be operated with three or more inputs dependent on the array size. An example is given in Figure 3.5 where a pixel is disabled if two out of the three inputs are high, permitting a 4x4 sub-array to be active (with one or two high DCR SPADs disabled), yet ensuring the rest of the array is disabled. The downside is the in-pixel decoding logic expands slightly (from a two input 'NAND' to a 3 or more input 'NAND'), but again this can only be used for small active arrays.

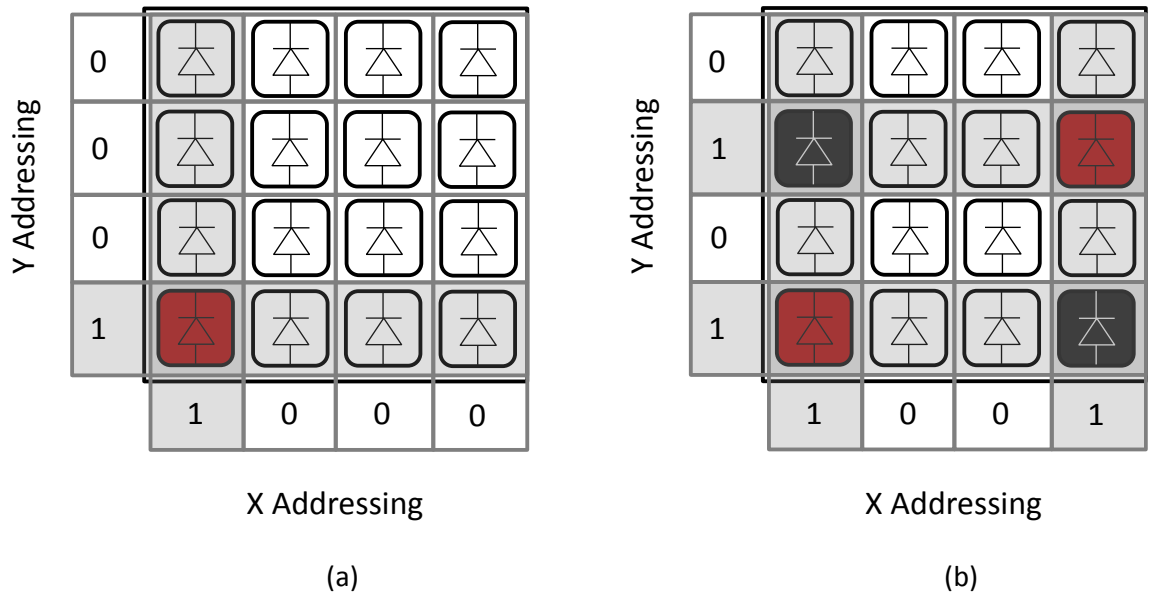


Figure 3.4. XY Addressing Logic Examples of SPAD Enable and Disable. (a) The high DCR SPAD highlighted in red is successfully disabled. (b) The high DCR SPADs in red are disabled but two working SPADs in black are incorrectly disabled.

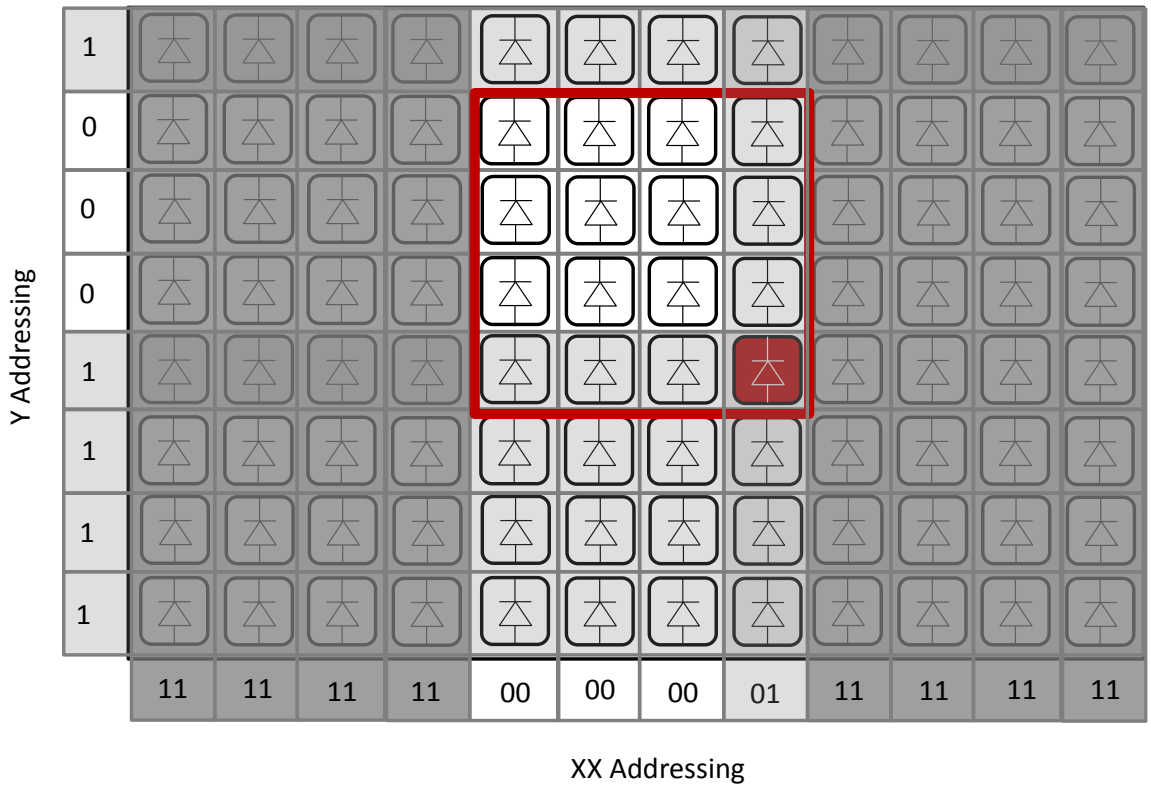


Figure 3.5. Extended XY Addressing scheme with two X-addresses and one Y-address for a SiPM requiring a small 4x4 sub-array (highlighted in the red rectangle) to be active whilst ensuring the rest of the array is deactivated.

For larger arrays, a static memory is the most reliable, as it can be programmed in an initialisation period before continuous operation. However, a static memory such as a 6T SRAM, latch or flip-flop requires a PMOS N-Well (NW) and has an area expense which would dominate a small pixel area and cause a loss of fill factor. A dynamic memory based only on NMOS devices, would need to be refreshed (like a DRAM memory array) to account for leakage, at least a few hundred times per second, translating to multiple refresh memory writes in exposures longer than a millisecond. If an analogue SPAD pixel is implemented with a dynamic memory, then the refresh cycles would cause significant temporal noise in the resulting image if not handled correctly. If a stacked technology is implemented then a static memory on the supporting logic layer is an ideal solution.

If the SPAD is connected with a poly resistor quench and recharge device then the SPAD cannot be disabled. Instead the SPAD output signal may be disabled or disconnected from the rest of either the pixel or system electronics. A dynamic memory was trialled for this purpose by Dr Eric Webster in [48] with a 4ms memory retention time. A static memory 6T CMOS SRAM was implemented, in parallel to this doctoral research, with a single initialisation period at device switch on without the need for memory refresh cycles (see [5] and Appendix 2).

In a SPAD-based image sensor, a transistor-based quench permits disabling operation where a poly-resistor does not. The benefits of a SPAD that uses a poly resistive quench must be weighed up against the inability to deactivate it. This disabling function is important for applications where external supplies are unavailable and the performance of an on-chip high voltage charge-pump based supply would be impaired by a high population of noisy SPADs. In the work described in this thesis, power consumption is not the primary concern as the SPAD high voltage is generated by an off-chip regulated supply and so high DCR SPADs are tolerated. In order to lessen any possible design complications with the array implementations developed in this thesis (described in chapters 4 to 6), SPAD enabling or disabling logic is avoided. However, the choice of SPAD is made ensuring that future revisions of this thesis work may have the facility to enable or disable the SPAD.

3.2.3. Directly Connected or Capacitively Coupled SPADs

The SPAD is interfaced to the subsequent logic in one of two ways: direct connection or capacitively coupling. In the former, interfacing the SPAD to CMOS logic is achieved by directly connecting the SPAD moving node to processing circuitry only if the SPAD's stable DC state and excess bias voltages are compatible with the CMOS logic. Otherwise in the latter, where the moving node has a DC bias at the high operating voltage (positive or negative), a metal-oxide-metal (MOM) capacitor is used to couple the SPAD to the rest of the circuit. There are a number of examples of coupling capacitors described in work by the CSS research group. Dr Richard Walker compares the different P-well to deep N-well options with MOM capacitors in [41]. As

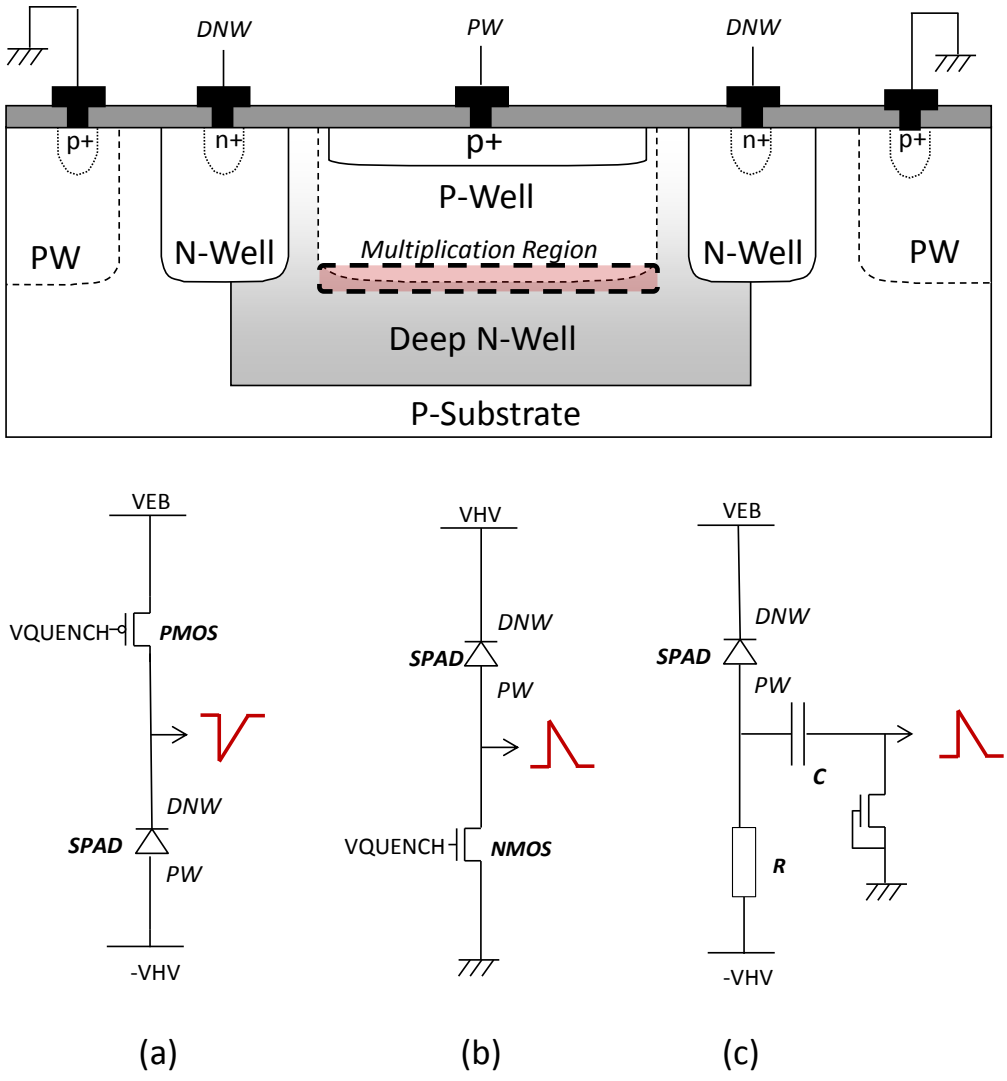


Figure 3.6. Upper part showing a PW to DNW enhancement SPAD layout cross section, the lower part showing three examples of PW to DNW SPAD coupling. Direct connection is shown in (a) negative drive and (b) positive drive and a capacitively coupled positive drive SPAD is shown in (c).

discussed in the later SPAD comparison section in this chapter, Dr Eric Webster implements a MOM decoupling capacitor in the deep N-well to P^- Substrate 'deep' SPAD pixel circuit [97]. Also a capacitively coupled 'deep' SPAD 32x32 SiPM array was implemented in this doctoral research in [5] and Appendix 2.

Figure 3.6 illustrates three possible methods of connecting a P-well (PW) to deep N-well (DNW) SPAD. In the upper part of the figure there is a layout cross section of the diode and in the lower part there are three circuit diagrams (a) – (c) connecting the same diode by different means. In Figure 3.6(a), in a negative drive configuration, the DNW cathode is the moving node, it is directly connected to the pixel logic as its idle state is at the excess bias (or overvoltage) supply 'VEB' and it pulses down to ground during a SPAD avalanche event. In the positive drive circuit in Figure 3.6(b), the PW anode is the moving node and is connected to the pixel circuit as the idle voltage is at ground. Conversely, in Figure 3.6(c) the SPAD anode idle state is at the negative high voltage supply and so is decoupled from the subsequent logic in a positive drive configuration. An NMOS pull down on the circuit side ensures the circuit side DC level is initialised and maintained. Certain SPAD constructions require AC coupling and these are highlighted later in this chapter in Table 3.1.

As illustrated in Figure 3.7(a), when implementing a MOM capacitor, a parasitic capacitance to ground is formed on both sides (C_{PP} and C_{PN}). On the SPAD side, the parasitic capacitance is in parallel with the SPAD moving node and will increase the total nodal capacitance, increasing the SPAD recharge time. On the circuit side, the parasitic has a bearing on the pixel circuit performance because it creates a capacitor divider circuit and decreases the amplitude of the SPAD signal. The voltage on the circuit side (V-) is given by the capacitor divider equation 3.1, noting no contribution of C_{PP}.

$$V_- = V_+ \cdot \frac{C_{MOM}}{C_{MOM} + C_{PN}}$$

(Eq. 3.1)

To ensure the greatest coupling of signal across the MOM capacitor, the parasitic on the circuit side (C_{PN}) must be minimised. If C_{PN} is not minimised, this will force the SPAD to be operated at a higher excess bias to ensure the SPAD signal triggers the processing electronics. A rise in excess bias increases both PDP and the DCR across the array, affecting the system signal to noise ratio (SNR), and impairing performance in low-light and possibly increasing SPAD-induced system noise. Also, the MOM capacitor and first stage amplifier or buffer must be designed to handle the decreased amplitude of the SPAD signal. To mitigate the signal loss, the MOM capacitor should be produced with a much greater parasitic on the V+ node rather than the V- node ensuring the maximisation of the V- signal. Figure 3.7 shows the proposed circuit and layout cross section of an optimal MOM coupling capacitor.

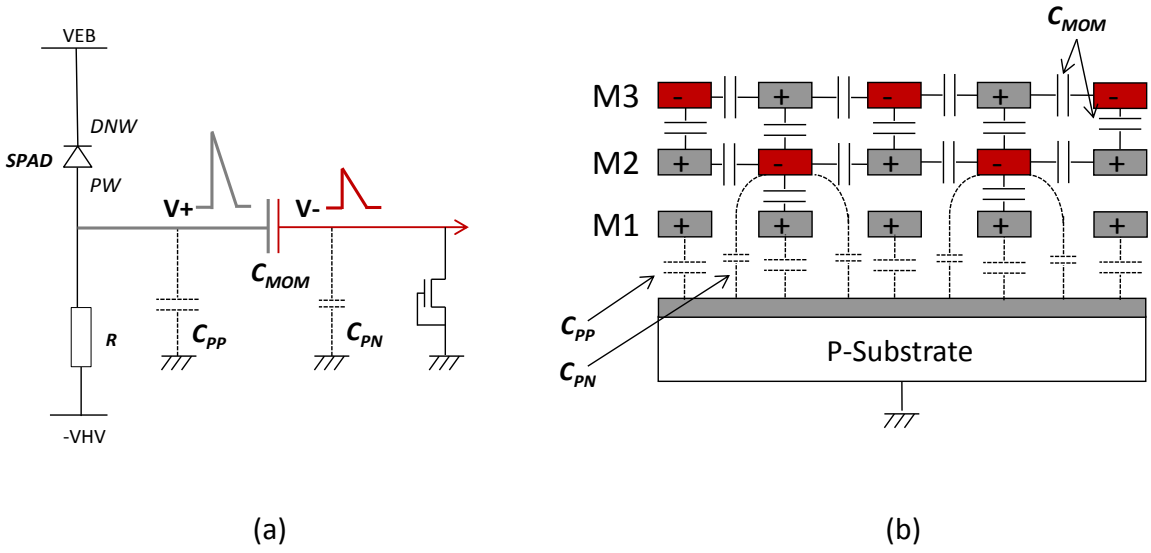


Figure 3.7. SPAD coupling capacitor design. (a) Circuit diagram showing parasitic capacitances C_{PP} and C_{PN}. (b) Layout cross section with minimisation of ground parasitic C_{PN} on circuit node V-.

The physical size and placement of the coupling capacitor places physical limits on the pixel scaling. In an FSI pixel, the device must be sized and located such that it does not impair the fill-factor. Furthermore, the closely spaced metal lines of the MOM device would cause an optical interference pattern to be created. If the capacitor is placed above or very near the active area of the diode, the sensitivity will be affected by the interference. As the MOM capacitor is reduced in size, greater variability in the capacitance contributes an additional source of variability to the SPAD recharge time. In a BSI pixel, of course the capacitor can be placed over the SPAD active area, or in a stacked process, it can be placed on the lower logic layer without the creation of optical interference.

In summary, a capacitively coupled SPAD may suffer from capacitor variability, optical performance impairment and increased dead time if not optimally designed. Choosing a SPAD that is directly connected does not suffer from these issues and further benefits from not having the design complications of implementing such a device.

3.2.4. Guard Ring Review

The previous sections deal with circuit implementations. This section is the first part considering the implementation of different diodes. Only CMOS compatible diode and guard ring structures are evaluated in this section. This section is not a complete or exhaustive literature review on guard ring structures. The interested reader is directed to the review paper in [68], or PhD theses [67], [98] for further information. Only well spacing design rule violations for these structures are considered; for the other design rules, as with any iterative technology development process, these are inevitably reconsidered, re-tested, rewritten and revised.

The guard ring design is an essential part of the SPAD construction. It serves two functions:

- To electrically connect the SPAD and physically isolate the high voltage wells from the surrounding substrate.
- To ensure that an even electric field is created across the intended photo-sensitive region of the diode.

There are three general methods to construct a guard ring: with an implanted diffusion well; by blocking the creation of the diffusion well to create areas of low doped silicon; or by trench isolation. The guard ring structure in the diffusion or low-doped guard rings, consists of three parts: inner guard ring, cathode well and outer guard ring. The SPAD inner guard ring serves one purpose: to ensure an even horizontal electric field is created across the multiplication region avoiding lateral edge breakdown between the diode active area and the guard ring itself. Incrementally reducing the inner guard ring width will eventually lead to this condition of edge breakdown producing an annular shaped active area with no sensitivity in the centre of the SPAD. The outer guard ring prevents lateral edge breakdown between the high voltage cathode well and the surrounding substrate. In some cases, the cathode well is not at the high operational voltage and so the outer guard ring is not required. The structure combining these three elements is referred to as the whole guard ring structure.

Each component part of the whole guard ring structure has a respective area penalty and strategies which decrease this area expense are considered in the context of a low pixel-pitch image-sensor pixel. There are two methods to reduce the area that are addressed in the latter part of this section:

- Minimisation of guard ring component widths.
- Well sharing.

3.2.4.1. Diffusion Guard Ring

There have been a number of devices published using a diffusion well guard ring. The first SPAD diffusion guard ring structure was described by Haitz et al. in 1963 to examine localised defects or 'micro-plasmas' in PN diodes via carrier multiplication [99]. The avalanche diode was implemented as a P+ to deep N-well, with a lower doped PW inner guard ring. For the interested reader, a concise history of the discovery of microplasmas through to modern day SPADs can be found in [69], and in greater detail in [67]. This structure has been implemented by Cova et al. [70], Kindt et al. [100] and more recently in CMOS by Rochas et al. [71] and in ST's process by Niclass and Henderson et al. in [101]. Figure 3.8 illustrates this structure. The P-well ring is lower doped than the P+ anode, preventing edge breakdown of the P+ implant, which would create an annular active region rather than the intended even circular multiplication region.

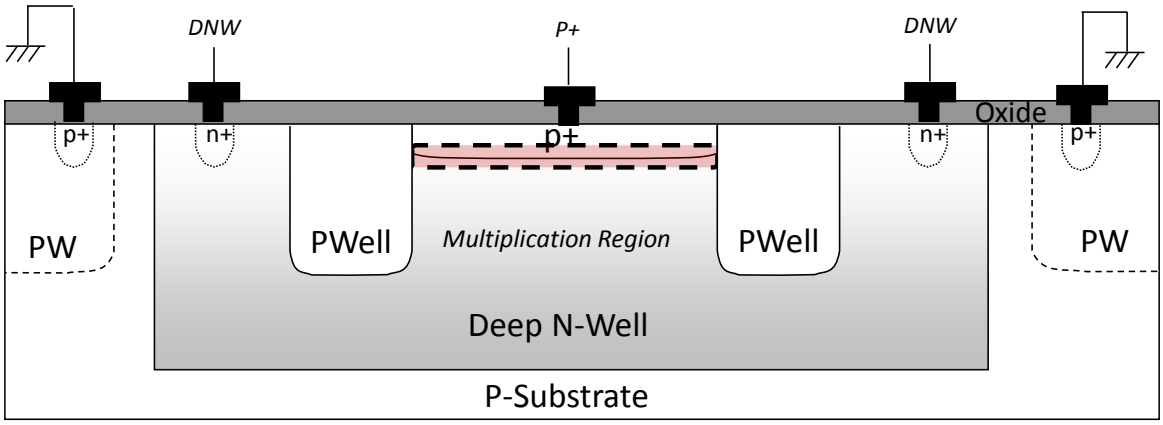


Figure 3.8. P-well diffusion guard ring.

3.2.4.2. Trench Isolation Guard Ring

A more compact guard ring structure was trialled by Finkelstein et al. replacing the PW guard ring with a shallow trench isolation (STI) etch in a 0.18µm process in [102], [103], shown in Figure 3.9. The aim of the trench was to significantly reduce the width of the guard ring in order to achieve a compact SPAD-based image sensor pixel. Furthermore, they report in further work the lowest jitter reported for a CMOS SPAD at 27.4ps FWHM [104]. However, due to the defects present at the STI-etch edge, the STI-bound SPADs suffer from high DCR in the megahertz region and high after-pulsing. This SPAD structure has the potential to be the most compact but is discounted for the proposed research applications due to the high dark noise.

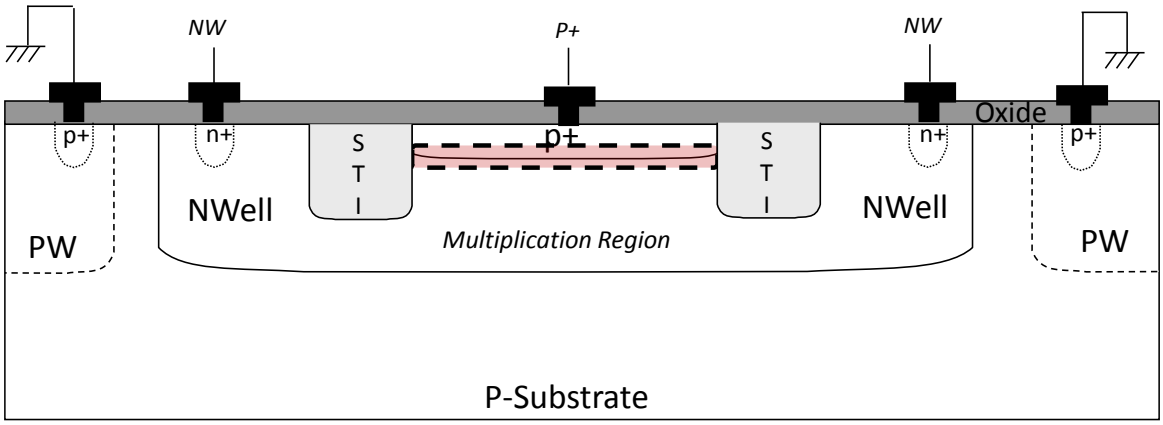


Figure 3.9. Finkelstein STI guard ring.

3.2.4.3. Low-Doped Guard Ring Structures

The first of the low-doped guard ring structures was trialled by Rochas et al. in [33]. A variant of the diffusion well guard ring, this structure is created by implementing a small space between the N-Well diffusion guard ring and the N-Well diffusion under the multiplication region. This gap lowers the doping at the edge of the diode reducing the peripheral electric field, whilst still having an electrical (although highly resistive) connection illustrated in Figure 3.10.

An alternative low-doped guard ring structure can also be created by increasing or enhancing the doping of one of the implants creating the multiplication region. This enhancement guard ring structure induces the high field region in the enhanced region, allowing a comparatively lower electric field at the edges of the diode, as seen in [105].

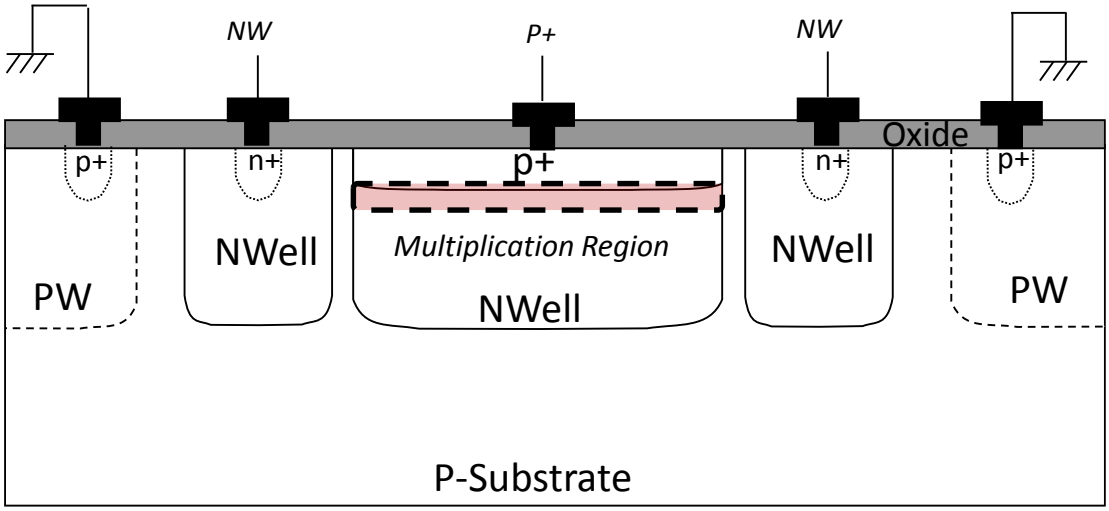


Figure 3.10. Low-doped guard ring structure with closely spaced N-Wells.

3.2.4.4. Virtual Guard Ring

An extension of the low-doped guard rings was conceived by research group member Dr Justin Richardson in his doctoral work, blocking both P and N-Well diffusions leaving only the epitaxial P- substrate ('epi') in both inner and outer guard rings [106]. This was termed the virtual guard ring structure as shown in Figure 3.11(a). The advantages of this structure are described in detail in Richardson's thesis and in [92] namely: reduced diffusion-dominated jitter tail (improved timing resolution over well-diffusion guard ring structures and non-substrate isolated devices), and no depletion region merging when scaling down the SPAD pitch and implementing this structure in a shared deep N-well. This concept was extended by Dr Eric Webster in the 'deep' SPAD shown in Figure 3.11(b), removing the PW anode and inner guard ring from Richardson's SPAD and creating the DNW to P-Substrate 'deep' SPAD [48], [67], [69], [97]. Extensive research was undertaken in those works to optimise the virtual guard ring structure in terms of DNW to NW overlap and the width of the epi outer ring.

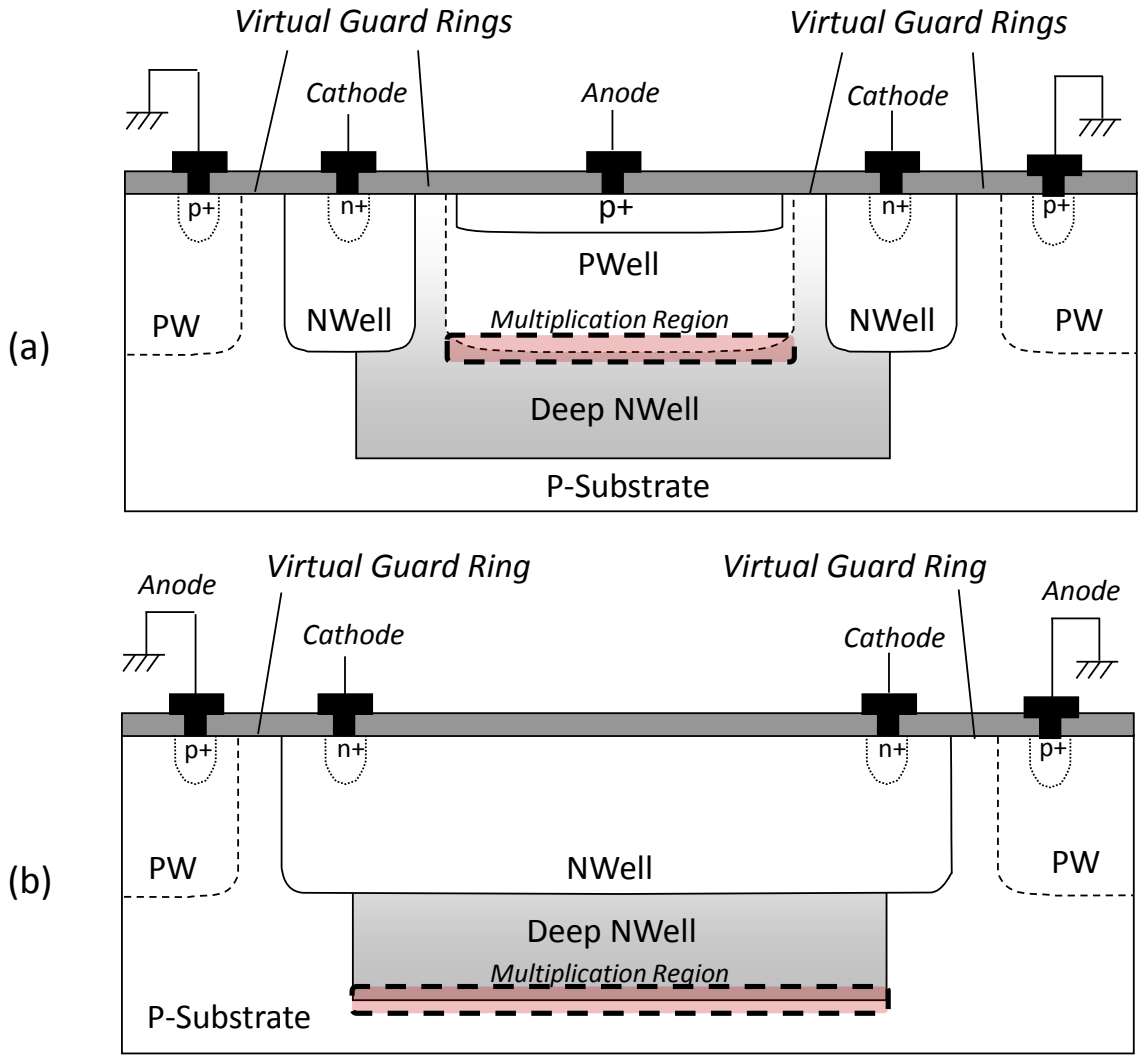


Figure 3.11. Virtual guard ring structures. (a) Richardson P-well to deep N-well SPAD. (b) Webster deep N-well to P- Substrate 'deep' SPAD. The scale is intentionally arbitrary.

3.2.5. Well Sharing

In a previous diode scaling study by Richardson et al. [107], the whole guard ring structure was maintained at the same radius (5.4μm) for sets of experiments increasing the active diameter. This study is revisited here, where Figure 3.12 demonstrates the area cost (as a percentage of the total area) of the guard ring versus the fill factor for the set of circular anode diameters tested (the pitch is calculated as the active area plus the guard ring diameter although not accounting for pixel to pixel hot well spacing rules). In the smallest device, the fill factor is 1.2% and the guard ring dominates with 77% of pixel area. This highlights the need for a strategy to achieve a compact and effective guard ring design to reclaim the pixel fill factor, in order to implement low pitch pixels.

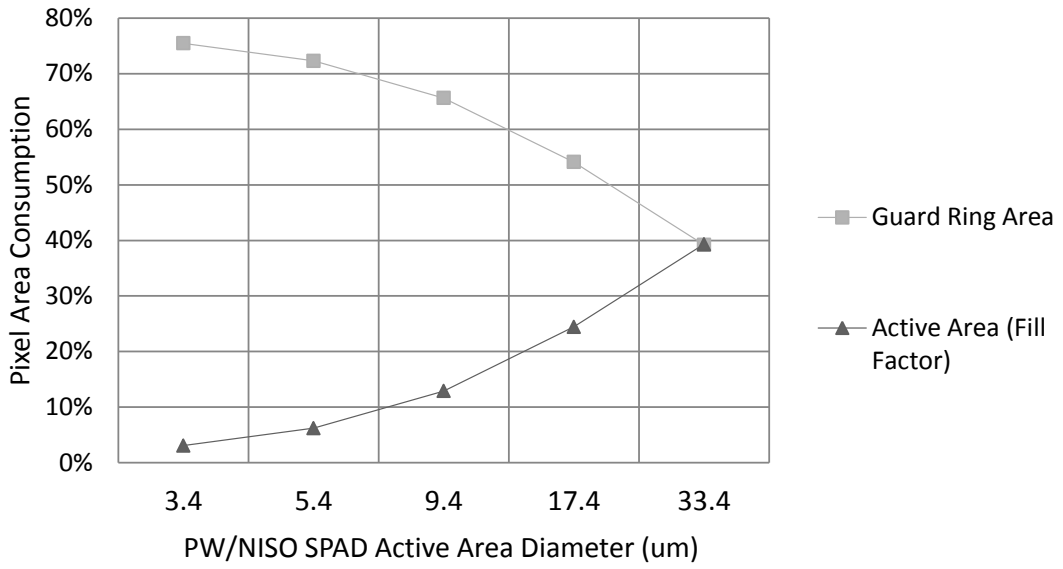


Figure 3.12. A graph of the percentage area consumption of the SPAD guard ring versus the active area using layout dimensions from the circular SPAD scaling study in [107]**.**

In Rochas' P+ to DNW structure [71], Richardson's negative drive PW to DNW [98] and Webster's 'Deep' SPAD structure [67] the DNW is the moving node. In Cova's N+ to P-Sub structure the N+ and NW guard ring comprise the moving node. The term for these SPAD constructions is non-substrate isolated SPADs. In all these cases the cathode NW of each SPAD in an array must be separated by well spacing design rules. Furthermore, no part of these SPAD constructions can be shared with another diode. This limits options for pixel pitch minimisation.

On the other hand, in Richardson's positive drive PW to DNW SPAD with virtual guard ring, and Niclass' P+/PW to DNW with diffusion guard ring [78], the P+ or PW is the moving node and the DNW is biased at the high operating voltage. In these cases, the SPAD is isolated from the substrate by the NW or DNW and such devices are termed substrate isolated SPADs. The primary advantage, in terms of pixel scalability, is this NW guard ring structure can be shared with other SPADs to create a well-sharing structure; each SPAD moving node is discrete and separated but the guard ring is shared. These are defined as shared well SPADs. There are two distinct methods of well sharing: global and local. The latter was conceived and developed as part of this research work.

3.2.5.1. Global Well Sharing

Global DNW sharing was proposed in the SPADNET project, as the method to achieve the highest fill factor SPAD SiPM macro-pixel. A Richardson PW to DNW SPAD is implemented where each PW anode is individually routed out and the DNW cathode is globally shared. In a monolithic image sensor context, in-pixel circuitry is of course required and may be placed beside the SPAD anode, if using an enhancement SPAD structure, in an isolated PW in the same DNW. This would have the advantage of not needing an outer guard ring structure and would save pixel area. A P+ to NW or DNW enhancement SPAD with three NMOS transistors in isolated PW was trialled in a global DNW by Henderson, Richardson et al. in ST's 90nm imaging process, creating the smallest SPAD pixel pitch yet attained [105]. It was successful as the P+/NW SPAD had a lower breakdown voltage than the transistor PW to DNW diode. However, the use of the PW as an anode denies the use of a PW for NMOS transistors using the same DNW, as the transistor PW would breakdown at the same voltage as the SPAD anode PW, without process modification. Furthermore, the use of PMOS transistors is prohibited in a global DNW scheme as the NW is biased at the high operating voltage. Otherwise, as shown in Figure 3.13, the circuitry must be placed at the array periphery. Like the individual routing of SPAD enable signals, this approach of individual SPAD nodes routed to the array edge is only scalable in one axis.

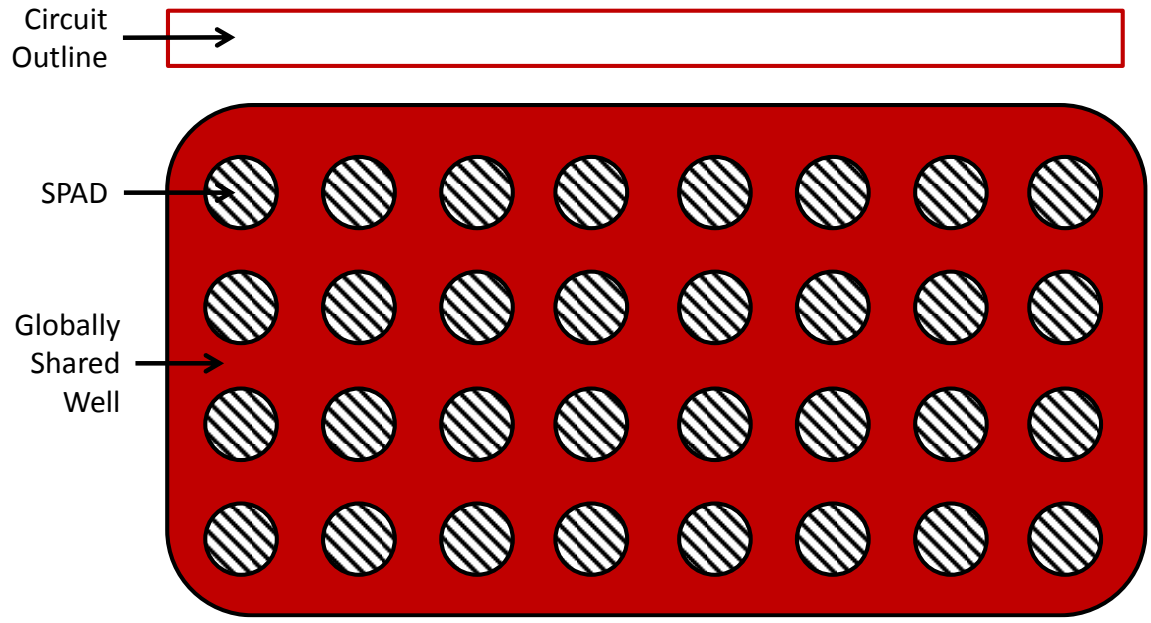


Figure 3.13. Simplified diagram of global well sharing layout. Electronics are at the periphery of the global well. Each diode is routed out individually.

3.2.5.2. Local Well Sharing

Local DNW sharing is proposed in this research as an effective compromise for an image sensor pixel, between the number of diodes shared in a well versus high fill factor. The concept is to separate the global well into separated local wells either as a group of four or in a grouping along an axis in a strip format. Figures 3.14 (b) – (d) describe the three options of local well sharing: 'quad', 'single strip', 'double strip'. With every break of the DNW, an outer guard ring is required which consumes pixel area and lowers the fill factor attainable. Figure 3.14(a) depicts a conventional separated diode pixel structure, where no well sharing is employed and the outer guard ring is required on all sides of the SPAD. In contrast, in the 2x2 or 'quad' local well sharing structure in Figure 3.14(b) and in the single strip structure in Figure 3.14(c), the outer guard ring is required only on two sides in each pixel. Extending this concept further to the dual strip SPAD structure in Figure 3.14(d), the area of the outer guard ring per pixel has been effectively reduced to a single side with a visibly greater fill factor than the separated diode structure.

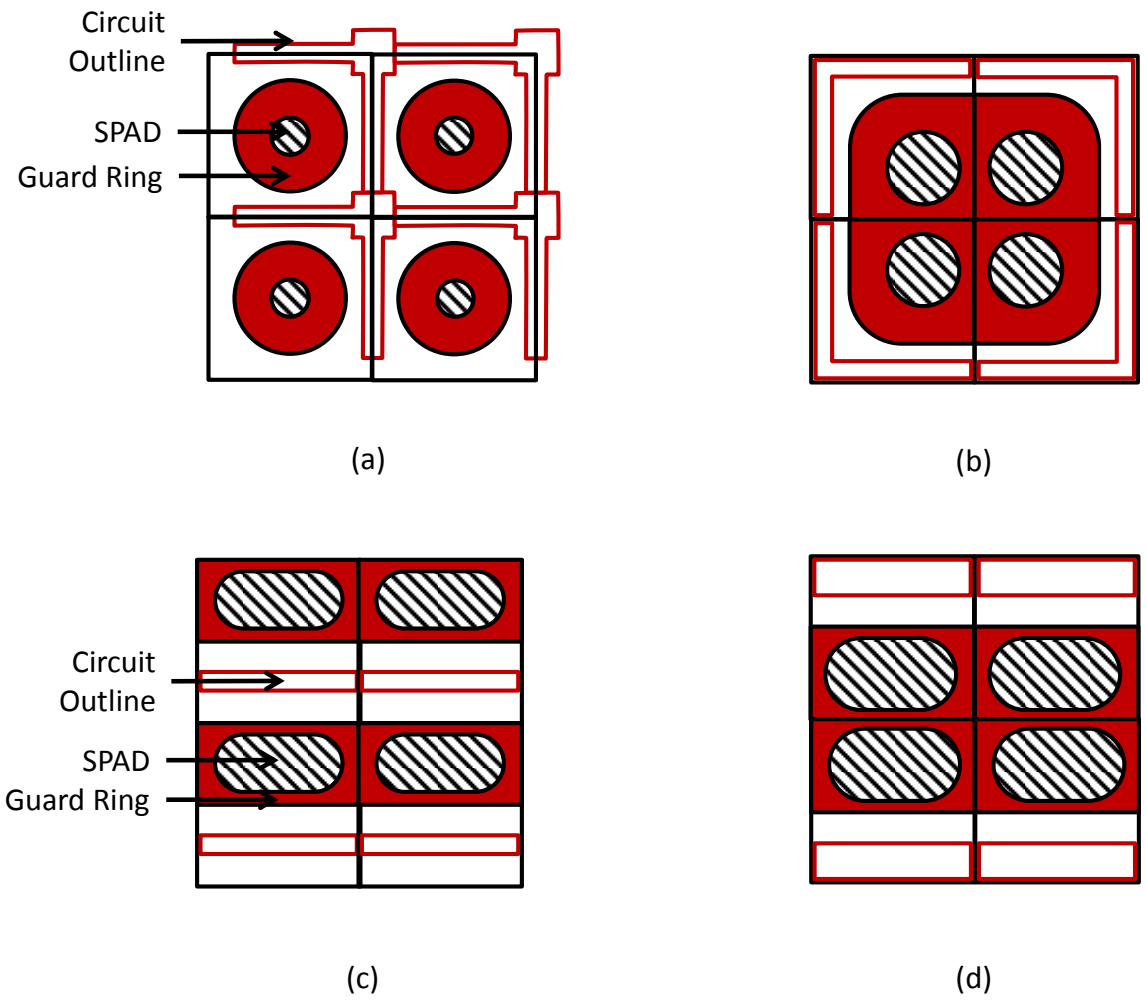


Figure 3.14. Top layout views of four SPAD pixels (a) Separated SPADs, no well sharing for comparison. (b) Quad local well sharing: four SPADs per well. (c) Single strip local well sharing. (d) Double strip local well sharing.

A twin structure (two anodes in a shared well) is intentionally omitted from this work as the arrangement had been conceived separately but not pursued by ST colleagues [108]. (The fill factor and pixel pitch obtainable with a twin structure would be approximately midway between what is obtainable with an isolated diode and quad local well sharing described below in Figure 3.16.)

Moreover, utilising the single or double strip format permits redesign of the SPAD anode. As illustrated in Figure 3.15, a circular anode would either create a non-square pixel (Figure 3.15(a)) or would have a redundant area of shared well (Figure 3.15(b)). However, as shown in Figure 3.15(c), in order to maximise the fill factor of the lowest pitch pixels, the circular anode may be elongated whilst keeping the circular ends to maintain a consistent electric field across the multiplication region of the SPAD. Extending this concept, as the pitch is increased the anode may be reshaped as a rectangle with rounded corners as shown in Figure 3.15(d) to maximise the fill factor. Furthermore, by breaking the DNW, it allows a PMOS NW to be placed in the pixel if required in the area marked 'Circuit Outline' at the cost of hot well spacing rules. Compared to substrate isolated diodes and global well structures, the downside to local well sharing is the loss of the regular arrangement or spatial uniformity in either or both horizontal and vertical directions. In an imager, this would create different modulation transfer functions (MTF) of the resulting image depending on the orientation of the sensor. Although not the basis of the idea, it is noted that the quad local well sharing structure is the SPAD pixel equivalent layout of the conventional 1.75T photodiode pixel.

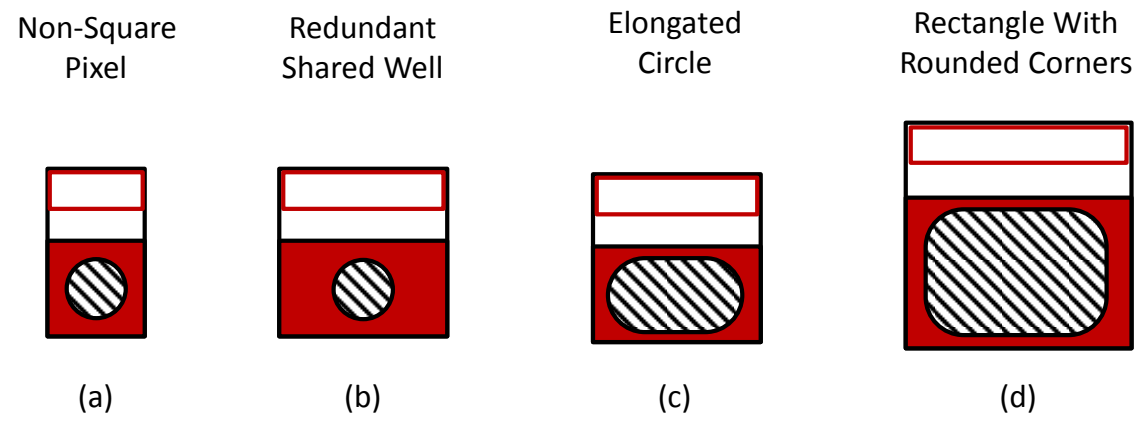


Figure 3.15. Single or double strip local well sharing anode design (a) Circular with non-square pixel (b) Circular with square pixel creates redundancy in the shared well (c) Elongated circular maximises fill factor for small pitch pixels. (d) Rectangle with rounded corners for larger pitch pixels.

First Order Optical Efficiency	Isolated	Quad	Single Strip	Dual Strip
X-axis	0.46	0.64*	0.87	0.87
Y-axis	0.46	0.64*	0.57*	0.70*

Table 3.1. First order optical efficiency in X and Y axes. Values normalised from a 20µm pitch example. * denotes the optical area is not centred and will lead to higher order spatial effects in the modulation transfer function of the sensor. The grey highlighted cells indicate where the values are equal.

Table 3.1 highlights the differences in optical efficiency (the ratio of photosensitive area to pixel area) in the different spatial X and Y axes of the sensor, for the different local well sharing approaches. The isolated and quad structures are equal in X and Y so create uniform spatial sampling and therefore equal MTF in both X and Y. Although the dual and single strip structures suffer from non-equal spatial sampling, they have higher optical efficiency or fill factor. The non-centred photo-sensitive regions will yield higher order spatial aliasing. For TOF applications this will create MTF patterns in the time-domain image although this is not considered a priority in this research. However, these effects will need compensated in applications that rely on even spatial sampling such as super-resolution microscopy. Appropriate microlensing could assist in mitigating these spatial aliasing effects.

In summary, the local well sharing scheme shares the area cost of the outer guard ring over a group of pixels in order to retain the advantage of higher fill factor that global well sharing brings. Contrary to global sharing, the local method allows both NMOS and PMOS transistors to be included in the pixel regardless of the SPAD construction.

3.2.5.3. Scaling of Well Sharing Structures

This section details a model that evaluates the trade-off between pixel-pitch and optical fill factor for the global well sharing structure, the three local well sharing structures, and the separated diode. Non-substrate isolated diodes are evaluated in the following sections of this chapter. In this modelling, the global well sharing structure also represents the pixel-pitch to fill-factor ratio of a stacked SPAD pixel with a photo-sensitive SPAD top layer and a bottom pixel logic layer.

The well sharing examples are applicable to both diffusion and virtual guard ring SPAD structures. The example presented here is for an NMOS only pixel. However, the conclusions and trends in this example still hold if PMOS devices are added with the important note that the fill factor would require recalculation, because adding a PMOS NW would lower the maximum achievable fill factor dependent on the size and placement of the NW.

To evaluate the pixel pitch of each of the structures, the following equations have been devised to model the geometrical spacing requirement of the constituent components of the pixel: SPAD anode, inner and outer guard rings, shared cathode and well spacing design rules. The pixel-pitches, of the five SPAD structures, are modelled with the following geometrical pixel pitch equations, where D = Active Diameter, I = Inner Guard Ring Width, T = OuTer Guard Ring Width, S= Shared Well Width, H_{NW} = Hot NW Spacing:

- Global Well Sharing / Stacked Pixel: $PP = D + 2.I + S$
- Double Strip Local Well Sharing: $PP = D + 2.I + T + {}^{S}/_{2} + {}^{H_{NW}}/_{2}$
- Single Strip Local Well Sharing: $PP = D + 2.I + T + {}^{S}/_{2} + H_{NW}$
- Quad Local Well Sharing: $PP = D + 2.I + T + {}^{S}/_{2} + H_{NW}$
- Separated Diodes: $PP = D + 2.I + 2.T + H_{NW}$

These modelled equations are graphed for comparison in Figure 3.16 showing the fill factor attainable at the pixel pitch for the five different SPAD non-sharing and sharing constructions. The kinks in the graphed lines mark the onset of rectangular anodes with rounded corners. The isolated diode line displays the limit of separated diffusion and enhancement SPAD structures using a P-type anode to DNW cathode using the design rules in the process design kit (PDK) design rule manual (DRM). The global well sharing approach shows the fill factor limit of the technology and shows the clear advantage that a stacked pixel would gain by removing the transistors out of the pixel, and into the lower logic layer. The three local well sharing methods proposed in this chapter show improved fill factor efficiency over the separated diode pixel using the quad, single strip and the double strip. The double strip local well sharing offers the highest fill factor at the cost of different MTFs in horizontal and vertical axes.

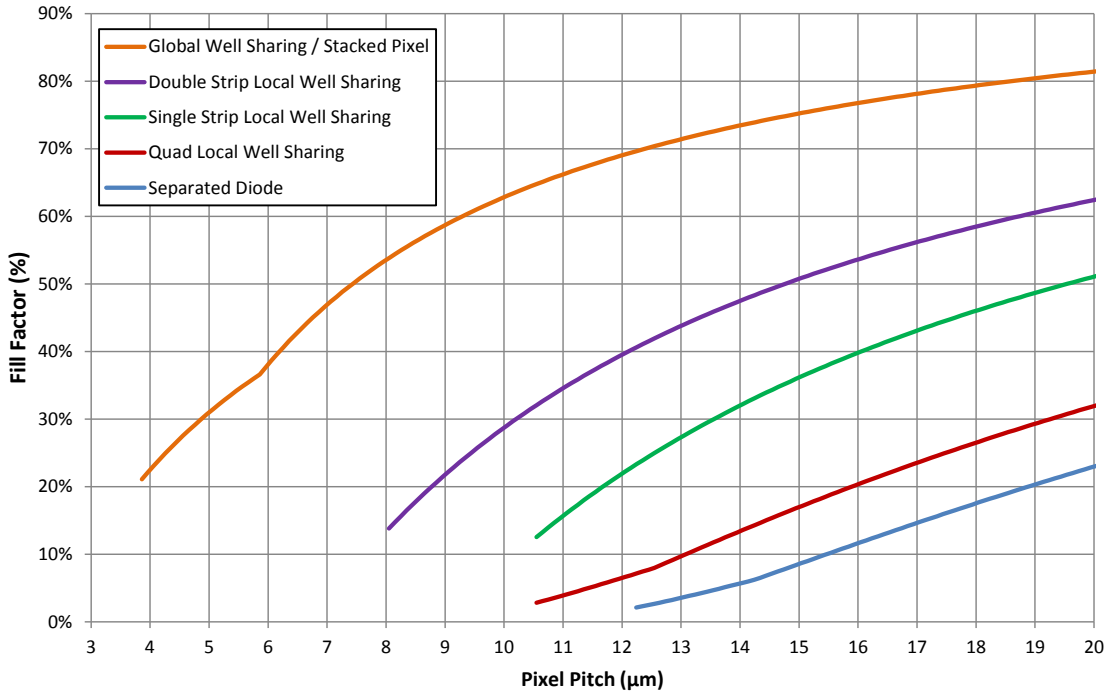


Figure 3.16. Pixel pitch versus fill factor for an isolated SPAD, three local well sharing approaches and the maximum attainable global well sharing approach with no transistors (the latter also emulates a stacked SPAD image sensor pixel).

The modelling shown is based on the PDK DRM. In the work presented in Chapters 4 to 6, design rules are re-evaluated to achieve higher fill factors than shown in the model. The following assumptions are made:

- The minimum space between pixels is dominated by the 5µm 'Hot NW' spacing rule. This space provides a 2.5µm width PW diffusion strip per pixel to place the NMOS transistors consisting of:
    - Space from transistor PW to outer guard ring (epi width): 1µm
    - NMOS transistor PW width: 1.5µm
- The anode is incremented as a circular shape (or elongated circular shape) from a minimum radius of 1µm. At greater than 2µm radius, the anode is reshaped as a rectangle with 1µm radius circular corners.
- The pixel transistors are NMOS only. The addition of PMOS NW decreases the optical fill factor.
- Inner guard ring (diffusion well or epi width) is set as 0.62µm
- Outer guard ring (epi width) is set at 1µm
- Shared side diffusion well width is set as 0.62µm
- Outer guard ring diffusion well width is set as 2µm, which consists of:
    - DNW to NW overlap: 0.4µm
    - NW outer width: 1.6µm

3.2.6. Substrate Isolation or Non-Substrate Isolation

This section compares substrate isolated or non-substrate isolated diodes, in terms of pixel scaling and PDE. The diffusion guard ring substrate isolated diodes with shallow implant junctions and multiplication regions near the silicon surface, have limited PDP at longer wavelengths with reduced NIR performance less than or equal to 3% PDP at 850nm wavelength at 3V excess bias [101]. The Richardson PW to DNW structure is a substrate isolated diode but has a deeper multiplication region with the diode junction at the bottom of the PW diffusion with slightly improved NIR performance at 5% PDP at 850nm at 3V excess bias [92]. On the other hand, there are three examples of non-substrate isolated diodes using a DNW to P- substrate diode with deeply buried multiplication region [67], [109], [110] realising an improved NIR response with greater than 15% PDP at 850nm. This improved NIR response warrants consideration for a TOF image sensor pixel.

3.2.6.1. Scaling of 'deep' SPAD versus PW to DNW SPAD

The 130nm Webster 'deep' SPAD diode is considered in terms of pixel scaling against the 130nm PW to DNW Richardson structure with the five modelled approaches (separated diode, local well sharing and global well sharing). The 'deep' SPAD is a negative drive device (producing a negative going pulse), if the in-pixel

circuit is fully NMOS then a positive going pulse is required to 'switch on' the NMOS devices, thus requiring at least one PMOS device in a CMOS inverter. An NMOS-only inverter is an alternative but is immediately discounted due to the static bias current.

As illustrated in Figure 3.17(a) the pitch of the 'deep' SPAD pixel with no PMOS NW is determined solely by the hot NW spacing rule and the diameter of the SPAD active region. Webster proposes in his thesis that this spacing should be defined as the spacing between the deepest NW's i.e. the 'Hot DNW to Hot DNW' rule as demonstrated in Figure 3.17(b) [67]. In evaluating the possible pitches of the 'deep' SPAD pixels, the DNW spacing rule is referred as 'H_{DNW}'. Figure 3.17(c) shows the pitch of the 'deep' SPAD pixel with a PMOS transistor calculated diagonally across the pixel using the diagonal width of the PMOS NW 'P'. Webster alternatively proposes the use of SPAD DNW to PMOS NW spacing rule [67], referred to as 'H_{DNW_NW}'. The modelled pixel pitch equations are as follows:

- 'Deep' SPAD – NMOS Only – NW Rule: $PP = D + H_{NW}$
- 'Deep' SPAD – NMOS Only – DNW Rule: $PP = D + H_{DNW}$
- 'Deep' SPAD – CMOS – DNW to NW Rule: $PP = D + \sqrt{\frac{(H_{DNW\_NW} + P)^2}{2}}$

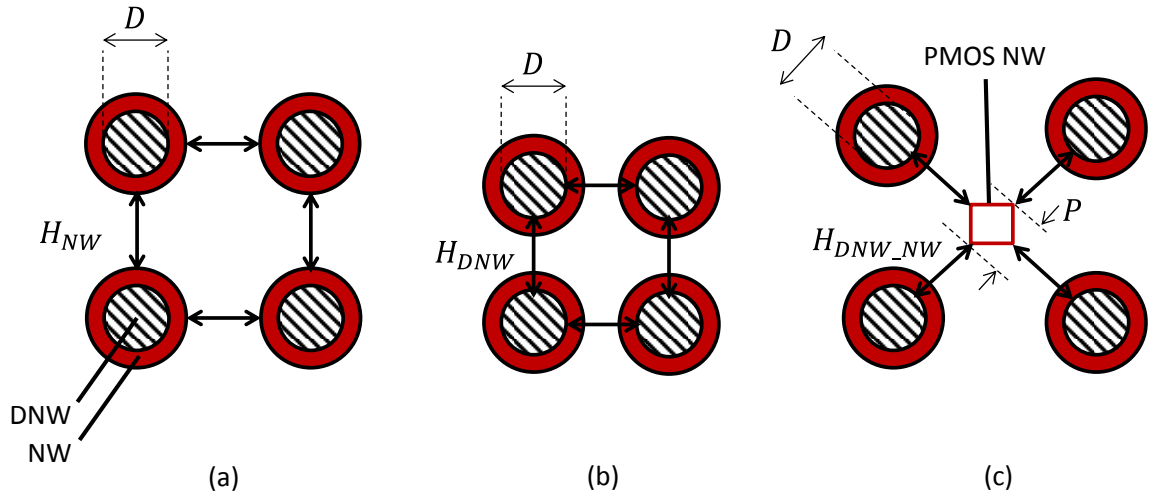


Figure 3.17. 'Deep' SPAD spacing rules. (a) Hot NW, (b) Hot Deep NW and (c) Hot SPAD Deep NW to PMOS NW.

Figure 3.18 displays the modelled pixel pitch against fill factor obtainable for the three 'deep' SPAD pixel arrangements and included for comparison are the global well sharing structure, the double strip local well sharing structure and the separated diode structure from the previous model. Although attaining a high fill factor comparable to the double strip well sharing approach, the NMOS only 'deep' SPAD pixels are not

considered practical for an imager pixel without a CMOS inverter. Instead, the yellow CMOS 'deep' SPAD line is considered the best achievable.

The published non-substrate isolated diodes have higher measured PDP (equivalent of intrinsic quantum efficiency (IQE)) in the red to NIR wavelengths due to the increased collection efficiency for deep generated minority carriers than the substrate isolated diodes. The PDE may be taken into account (product of fill factor and PDP, equivalent of extrinsic QE (EQE)) at the wavelength of interest of the imaging system, where:

$$PDE(\lambda) = FF\ .\ PDP(\lambda)$$

(Eq 3.2)

Figure 3.19 presents a graph of the modelled results of PDE versus pixel pitch using 20% PDP for the three 'deep' SPAD pixels and 5% PDP for the three substrate isolated diode pixels previously modelled. As discussed, the two most promising candidates for image sensor implementation, due to the highest fill factor, are the double-strip local well sharing structure and the CMOS 'deep' SPAD pixel. From the graph for imaging at 850nm, the double-strip substrate isolated pixel (purple line) has higher PDE at pixel pitches below 13µm than the 'deep' SPAD pixel (yellow line). The 'Deep SPAD NMOS Only' (green and red lines) are kept for comparison if the NMOS inverter bias current is acceptable in a given application.

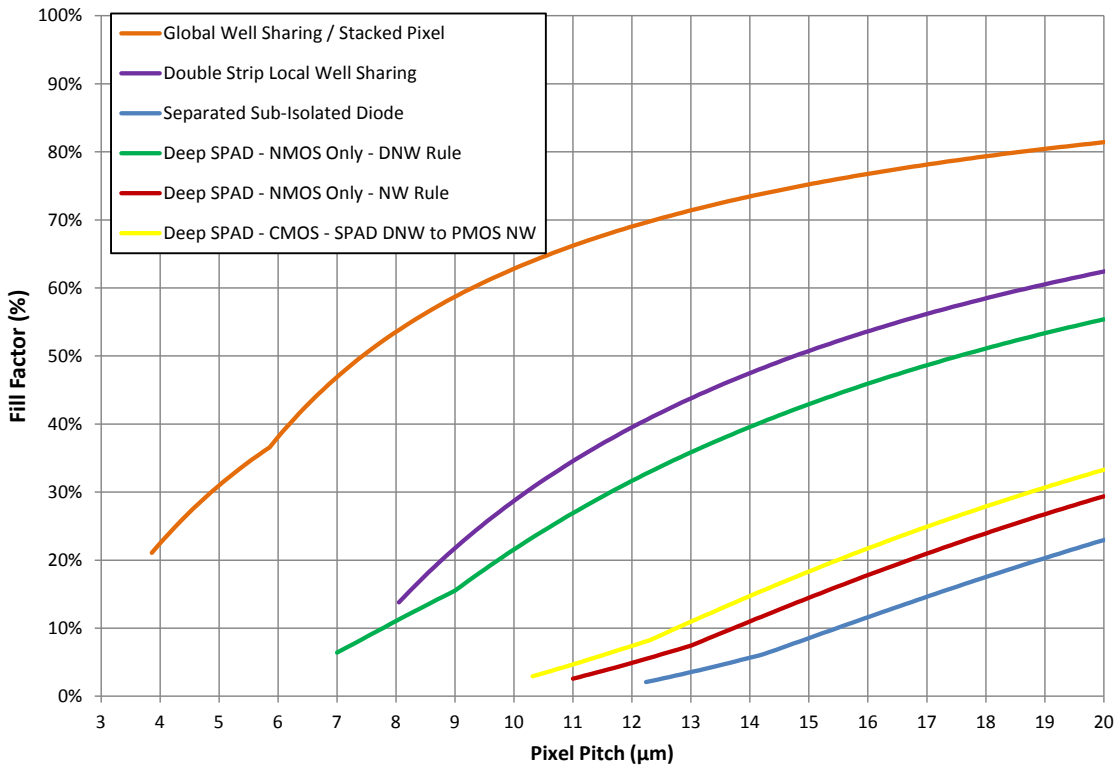


Figure 3.18. Fill factor versus pixel pitch for three 'deep' SPAD pixels, the global and double strip local well sharing approaches, and the separated diode pixel.

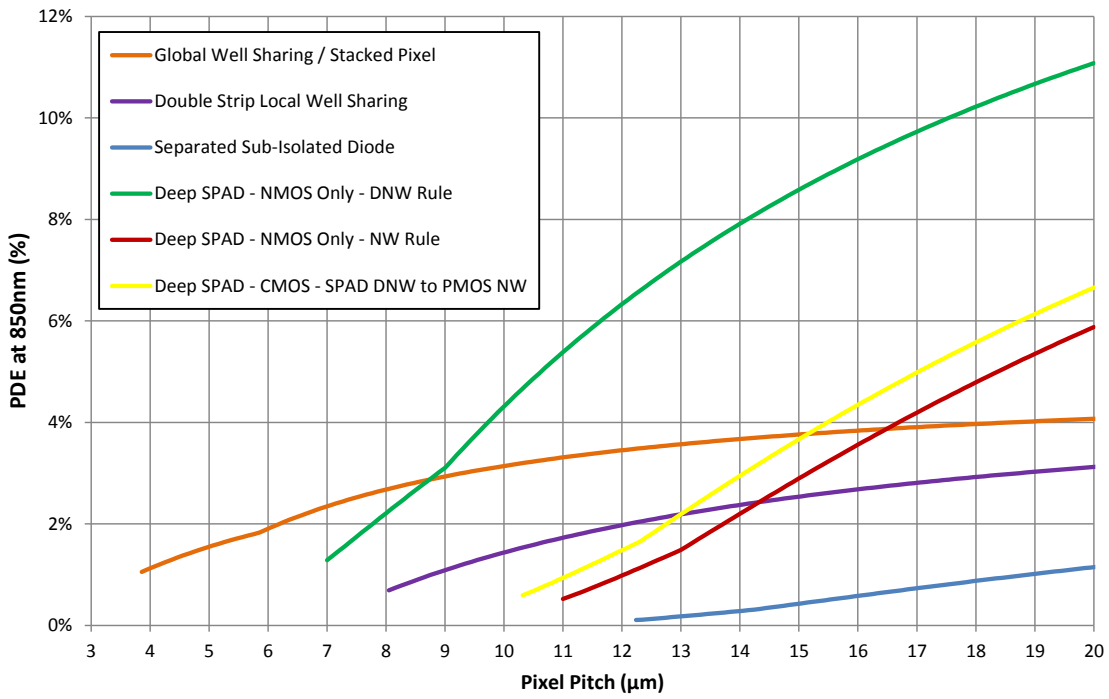


Figure 3.19. PDE at 850nm versus pixel pitch for three 'deep' SPAD pixels, the global and double strip local well sharing approaches, and the separated diode pixel.

3.2.7. Supplying the High Voltage

Supplying the operating high voltage for the SPAD is a design concern. For a prototype chip, the high operating voltage and all other supplies and biases are supplied by off-chip regulators. However, in a commercial SPAD vision sensor product, limited power connections are available (often only 3.3V and ground) compelling the designer to include on-chip regulators. In this case, an on-chip boost converter must be designed to generate the requisite SPAD high voltage. The design of the boost circuit must use the deep N-Well for high voltage P-well isolation. This imposes a constraint on the SPAD choice: the operating voltage (the breakdown voltage plus the excess bias) of the SPAD must be less than the breakdown voltage of the deep N-well of the boost circuit without process modification. In the 130nm process technology used in this research, the DNW to P-substrate breakdown voltage (of a correctly designed isolated-well structure) is approximately 20V.

3.2.8. SPAD Choice for Time Resolved Imaging Pixel

This section summarises the different possible diodes and their respective implementation requirements in order to choose the pixel for a time resolved SPAD-based image sensor. As discussed throughout section 3.3, the ideal SPAD will have the following characteristics:

- Positive going pulse for NMOS-only pixel electronics.
- Transistor based PQPR for possible inclusion of enable and disable logic, preferably NMOS PQPR for NMOS-only pixel.
- Direct, non-capacitively coupled connection.
- Junction breakdown voltage lower than the highest breakdown voltage in the fabrication process (<20V).
- High PDP and PDE at 850nm.
- Low temporal jitter at 850nm.
- Low DCR.
- Amenability to well-sharing techniques.

Table 3.2 lists the SPADs discussed and referenced in this section in relation to these performance or implementation parameters. The data in the table have a caveat that PDP and jitter are problematic to compare as different excess biases are used in each of the publications. The connection and PQPR columns are not from the publications themselves, but are evaluations if the diodes were fully integrated on-chip, as many of the works listed present isolated structures fabricated with off-chip PQPR and readout circuitry. The standard CMOS column denotes if the front end of the line requires custom implants or masks for the SPAD fabrication.

It is noted here that the increased reported PDP for NIR wavelengths of non-substrate isolated diodes is due to a combination of the detection of both drift and diffusion minority carriers. The slower diffusing carriers appear as tails in the timing response. On the other hand, substrate-isolated diodes will have greater isolation of the multiplication region, from diffusion minority carriers, suppressing the jitter tail. This is critical for temporal imaging applications. A qualitative observation has been made by the author on the prominence of a jitter tail in the published histograms of timing response, in the final table column.

The Richardson PW to DNW structure is highlighted and selected as it meets all the listed implementation characteristics over the other diode structures. In comparison to the other substrate isolated diodes it has the highest PDP at 850nm although higher temporal jitter.

Author and Diode	Reference	Drive	Connection	PQPR	Breakdown Voltage (V)	PDP @ 850nm (%)	Substrate Isolated & Well Sharing	Jitter FWHM (ps)	Prominent Jitter Tail	Compactness	Standard CMOS
Richardson et al. PW to DNW	[106]	+	DC	NMOS	13.6	5	Y	184	N	+	Y
		+	AC	R	13.6	5	Y	Not reported	N	+	Y
		-	DC	PMOS	13.6	5	N	Not reported	N	-	Y
Richardson et al.P+/PSTI to DNW	[106]	+	Dc	NMOS	17.9	3	Y	237	Y	+	Y
Richardson et al. P+/PSTI to NW/DNW	[106]	+	DC	NMOS	12.4	2.5	Y	183	N	+	Y
Niclass et al. P+/PW to DNW	[101]	+	DC	NMOS	9.7	3	Y	144	Y	+	Y
Ghioni et al. P+ to N-Sub	[111]	+	AC	R	>33.8	15	N	35	N	-	N
Lacaita et al. N+ to P^-Sub	[112]	-	AC	R	20	Not reported	N	45	N	-	N
Villa et al. P+ to DNW	[87]	+	DC	NMOS	25	2	Y	56	N	+	Y
Finkelstein et al. P+ to NW	[113]	-	DC	PMOS	10	Not reported	N	Not reported	Not reported	-	Y
Veeperran et al. DNW to P-Sub	[110]	-	AC	R	23	15	N	90	Y	+	N
Webster et al. "Deep" SPAD DNW to P^- Sub	[97]	-	AC	R	20.2	20	N	49	Y	+	N

Table 3.2. CMOS SPAD choices for a time resolved SPAD-based image sensor. AC = AC Coupled, DC = Directly DC Connected.

3.3. Pixel Classification

This section seeks to categorise the myriad CMOS SPAD pixel designs in the literature. These are reviewed to determine an apppropriate choice for a SPAD-based image sensor architecture scalable to megapixel arrays. In CMOS SPAD pixel design, as in CIS pixel design, there is a trade-off and compromise among three factors: pixel pitch, fill factor and functionality. Amongst other factors, the large pitch and low fill factor of previous SPAD pixels has limited the uptake of the technology in applications such as scientific imaging and 3D vision that require both high spatial resolution and high sensitivity.

Image sensors have a different set of design constraints to single-point or line sensors. An ideal imager will have high frame-rate, low power consumption and maximal array size. An ideal pixel will, of course, have appropriate pitch, 100% fill factor, 100% QE or PDP at the range of wavelengths of interest, 0% QE at wavelengths out of interest, no noise generation or noise sensing, and minimal power consumption. In all image sensor technologies, the optical design of the sensor is paramount; achieving 100% fill factor at a minimal pitch (allowed by technology constraints and design rules) is the ideal goal but this gets traded off against the number of transistors in the pixel – more transistors providing more in-pixel functionality yet often to the detriment of optical efficiency. Micro-lensing may be used to reclaim the loss of fill-factor yet this does not redeem an inefficient design [114]. On the other hand, not every CIS pixel can be micro-lensed, the pitch must be compatible with the wafer-level micro-lens manufacturing process. Alternatively, micro-lenses may be added on a per-die basis as a post-processing step, which would significantly increase the per-die cost.

In this comparative review the pixel is defined as the individual light sensitive area in the imaging array, circuitry placed outside the imaging array is discussed but is not included in the pixel fill-factor calculations. The transistor choice has a sizeable impact on the pitch and fill-factor. In the context of monolithic SPAD sensors, NMOS transistors may be placed in close-proximity to the diode high voltage guard ring well but PMOS devices placed inside an N-Well must be placed at a safe-distance from it to avoid the possibility of horizontal bipolar latch up by merging depletion regions. As previously discussed in this chapter, this transistor NW to SPAD NW spacing in the pixel array design must be observed. As 3D chip stacking technology becomes available, the flexibility will be granted to put some or all the in-pixel electronics on the lower logic support wafer (using a lower process node standard CMOS technology) and to have a fully customised imaging CMOS technology for the top photo-sensitive SPAD layer. This technology will offer up numerous opportunities for the design of SPAD arrays, alleviating many current constraints such as well spacing.

SPAD pixels are broadly classified into six different sets for this review, based on the type of in-pixel circuitry and the number of transistors. The categories are:

- SPAD only with no transistors
- Minimal transistors
- Active quench and recharge

} SiPM pixel architectures

- All digital circuitry
- Single bit memory circuitry
- Analogue circuitry

} Image sensor pixel architectures

The first three categories are the basis of high fill factor small arrays and the building blocks for small imaging arrays or DSiPM arrays with only quench/recharge and addressing circuitry. On the other hand, the latter three categories have the capability to capture, measure and store a photonic event and are the basis of high resolution SPAD-based image sensors with scalable pixel arrays.

3.3.1. SPAD Only Pixels

The maximum fill-factor FSI pixel would contain only the photo collection region and no pixel transistors, comparable to a CCD pixel. However, in such a SPAD-based pixel each diode must be individually connected and routed out to the edge of the array. High fill factor is achieved by means of diode global well sharing covered previously. As illustrated in Figure 3.20, the pixel array is only scalable to a certain array size determined by the width of the SPAD guard ring and the metal width and spacings of the process technology for connecting each SPAD cathode or anode. As a greater number of SPADs are routed out, more metal routing is required. In the worst case example Figure 3.20(c) the cathode metal routing impacts the optical performance of a significant proportion of the imaging array. This spatial scalability limit that this approach imposes makes it unsuitable for an image sensor implementation capable of scaling to very large array sizes in a monolithic process. Furthermore, the central array pixels (furthest from the readout edge) will have a greater systematic timing offset than the edge pixels, although this could be calibrated and compensated for. However, in a stacked process, the spatial limit is lifted and timing balance can be achieved by balanced routing in the logic layer.

The 'SPAD-only' pixel approach is ideally suited to creating large silicon area optical sensors consisting of multiple discrete photo sensitive sites which are aggregated or summed, either in analogue or digitally, to create a single point device or large pixel sensors. The terms Silicon Photo-Multiplier (SiPM) [95], digital SiPM (dSiPM) [115] and Multiple Pixel Photon Counters (MPPC) [116] are used to describe such devices. Recently Hamamatsu have released a MPPC device consisting of 300x300 (90,000 total) SPADs at 10µm diode pitch with a single analogue summed output [117].

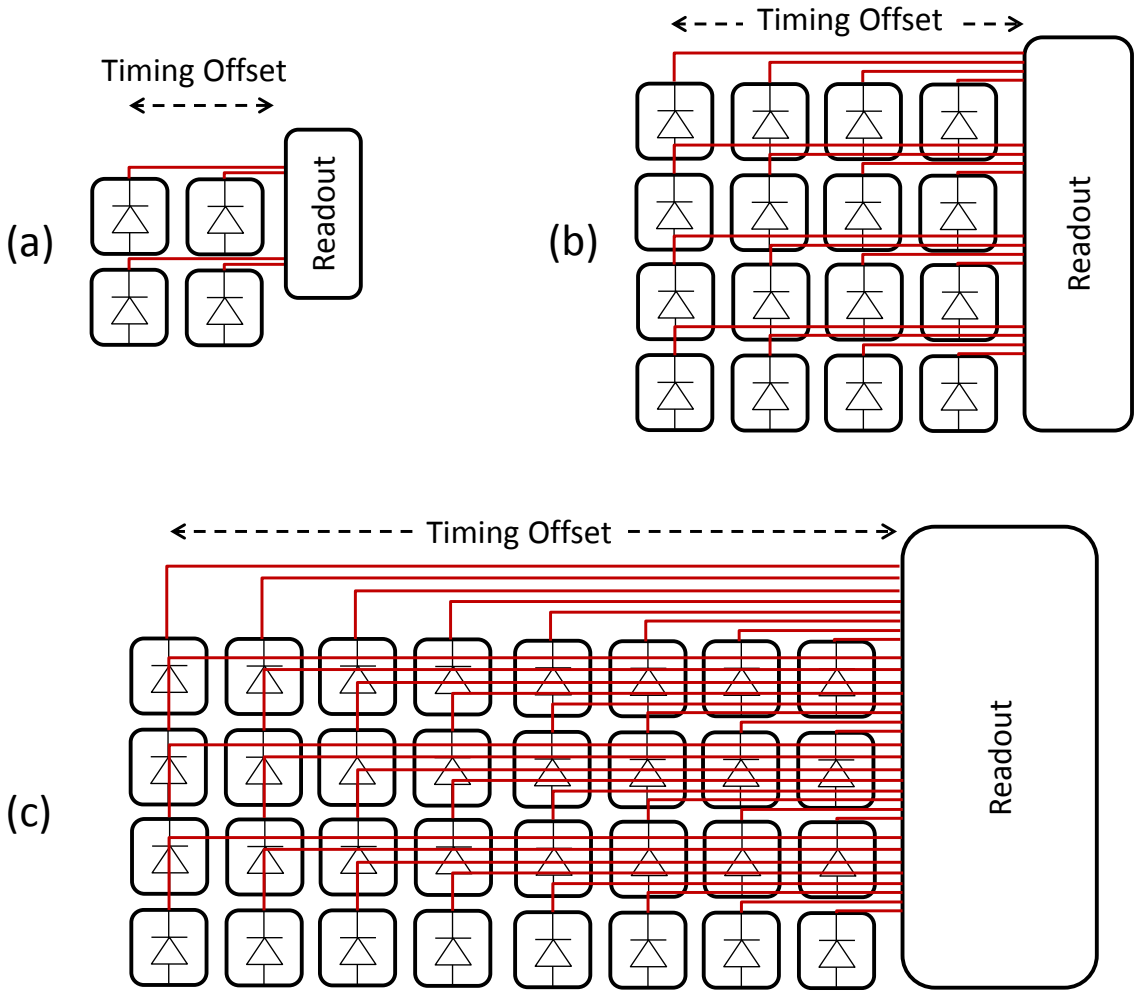


Figure 3.20. Spatial scalability limits of SPAD-only pixels with three examples of pixel routing and the effect on optical performance. (a) Routing has no impact. (b) Routing has impact for some pixels. (c) Routing significantly impairs majority of the array.

Another effective example of this technique is seen in the SPADNET TOF-PET sensors [41], [80], [82], [95], [118]. The first SPADNET sensor has 16x8 macro pixels, each macro pixel is a SiPM consisting of 720 SPADs with a diode active diameter of 16.27 μm in a hexagonal arrangement. Each macro pixel has a TDC and the sensor has on chip counters and digital signal processing (DSP) to indicate the time of arrival of photons from a scintillation event caused by incoming gamma radiation. As a stacked example, a 4x4 'SPAD only' macro pixel is shown in Figure 3.21 with pulse combining logic and TDC in the supporting lower logic layer. The limit of the stacked pixel array size is then imposed by the temporal pile-up in the routing and combining logic. Modelling of pile up and SPAD array design optimisation is presented in collaboration with research group colleague Salvatore Gnecchi in Appendix 2, [5], [10], [11] and in colleague Dr David Tyndall's research [77], [119], [120].

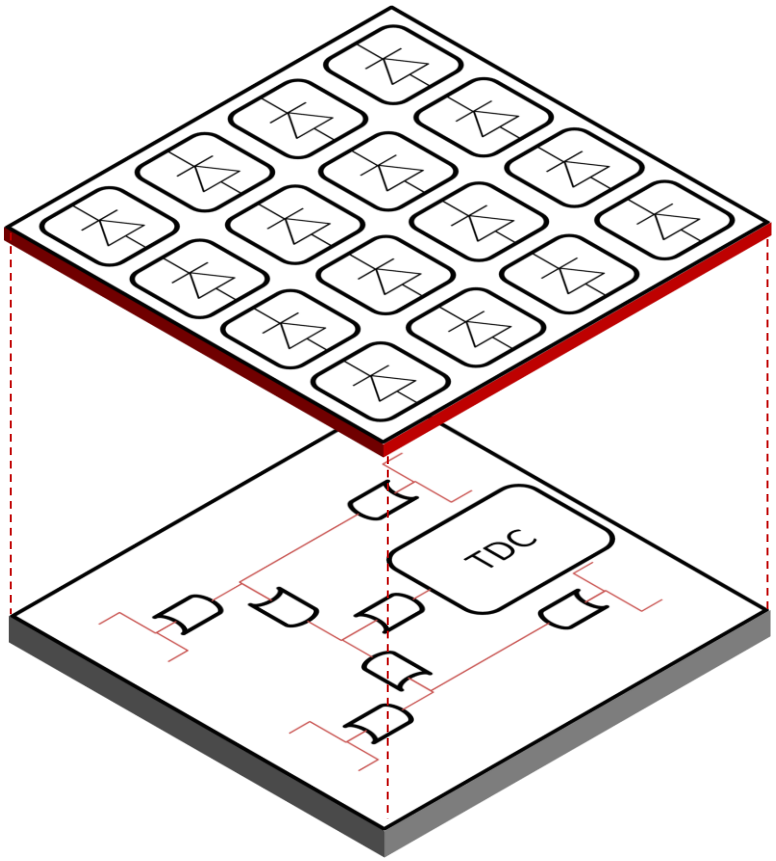


Figure 3.21. A SPAD-only macro pixel in chip stacking technology. The spatial scalability limit is lifted in comparison to a monolithic process. The logic layer contains timing balanced combiner logic and time to digital converter, or counting circuitry or both.

3.3.2. Minimal In-Pixel Circuitry

Instead of individual SPAD routings, there are a number of examples of adding a small number of pixel transistors and connecting SPAD pixels together on a column bus. Activating one row at a time, the column bus connects the buffered SPAD pulses to column parallel integrators or time converters. Figure 3.22 demonstrates three examples from the literature. Figure 3.22(a) shows a 5T solution, created by Niclass et al. in 2005, placing a PMOS quench, a CMOS inverter and a transmission gate in a 58µm pixel connecting a 32x32 array to a single channel TDC in 0.35µm technology [121]. Henderson et al. created the lowest pitch SPAD pixel to date (5µm) in ST's 90nm CMOS and used a NMOS-only approach employing a 3T pixel (quench, source follower and read select) using the circuit in Figure 3.22(b) in a 3x3 array [105]. Niclass et al. presented a 128x128 array of 25µm pixels with 32 column parallel TDCs to create a DTOF imaging array with the pixel design shown in Figure 3.22(c). An asynchronous event-driven readout was implemented per four columns, correlating the processed TDC events with positional data (from one of four pixels), and allowing all pixels in a row to be active and incrementally building a TOF image line by line [122].

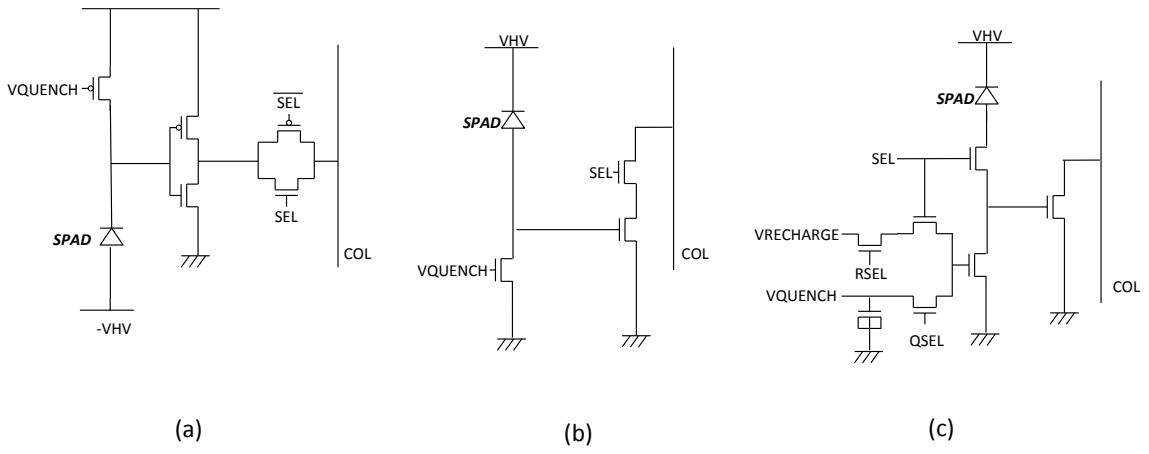


Figure 3.22. Examples of minimal in-pixel circuitry: (a) Niclass et al. 5T 58µm pixel [121] **(b) Henderson et al. 3T 5µm pixel** [105] **(c) Niclass et al. 7T 25µm pixel from 128x128 array** [122]**.**

A 16x16 SiPM architecture employing the 'Deep' SPAD was published by Dr Eric Webster using this column bus format with 11.6µm pitch at 21.6% fill factor [95]. Also a 32x32 SiPM architecture is described by Dr David Tyndall in [119], [120]. This column bus format is only applicable to systems and imaging applications where only a single row or small number of pixels is active at one time. The column bus could be shared if the frequency of events on the bus was suitably low. Yet, there is a downside to this approach: the SPAD dead-time after each photon arrival holds up the column bus disallowing subsequent events from other pixels during this period – causing a 'pile-up' distortion in the output routing channel [77]. This channel pile up effect is proportional to the number of pixels simultaneously connected to any output bus and worsens with increasing ambient light incident on the sensor, making the channel sharing critical to imaging system performance for applications in uncontrolled light environments.

Both the minimal circuitry and the SPAD-only approaches have been successfully implemented by Niclass et al. in the DTOF Toyota LIDAR imaging system. A number of line sensor pixel variants have been described by the Toyota team in [35], [78], [123], [124]. The latter three publications describe a 'minimal' transistor approach, but in the former [35] the 'SPAD only' approach is employed, in which all the pixel circuitry is removed to edge of the array to attain the highest recorded fill-factor for a SPAD pixel at 70%, with a square with rounded corners shape at 25µm pitch. Each pixel is routed onto both a column and row wise bus. The column channel is fed into parallel TDCs and row-wise onto event driven positional decoders. Ambient rejection logic, time gating and an optical notch filter are used to mitigate time converter pile-up and to lessen the DTOF processing required. Yet, as pixels are still connected together on a bus, this architecture still suffers from channel pile-up distortion if scaled to a larger array. This column parallel bus-sharing suits a line sensor or scanner-based system with many columns and few rows but is not suitable for scalable image sensor pixel design.

3.3.3. Active Quench and Recharge

An active quench and recharge circuit (or commonly known as 'active quench' or AQAR) is a monostable circuit with different paths for quench, and recharge of the SPAD moving node. The quench circuit pulls the SPAD into a disabled state for a short duration at the onset of each avalanche and the recharge circuit re-activates the SPAD ready to receive the next photon or dark event. Figure 3.23 illustrates a simplified schematic and timing diagram of the active quench and recharge process. The inverters in Figure 3.23(a) 'τ_Q' and 'τ_R' represent the delays (whether monostable or current starved inverter) required to implement the active quench and recharge functions. The 'τ_Q' delay is minimised and the 'τ_R' delay sets the balance between a minimisation of after pulsing and maximisation of count rate.

The active quench has been applied in both a SiPM and a small imaging array. Phillips designed a DSiPM macro-pixel of 64x32 SPADs for PET with each SPAD sub-pixel with 52x30µm pitch at an impressive 50% fill factor reading out to a single channel on-chip TDC [83]. Field et al. recently proposed a 64x64 SPAD-based TCSPC image sensor with active quench for FLIM [125], [126]. The 48µm pixel had low 0.77% fill factor due to the active quench circuitry and conservative layout. Using a DSiPM type array, each SPAD node is individually routed out to its own TDC at the edge of the array. This sensor architecture is restricted by the same spatial scalability limit as the 'SPAD only' devices. Eisele et al. [89] trialled a SPAD in ST's 130nm technology with active quench and recharge circuitry, attaining the highest count rate reported for SPADs at 185MHz. The pixel pitch is not reported in this work, but is estimated by the author at 22µm x 28µm with an 8.2% fill factor. Finkelstein et al. trialled an active quench in a 0.18µm process at 180µm x 96µm with an estimated 0.5% fill factor. The area cost, in all four examples of the active quench and recharge, is high with a number of both NMOS and PMOS transistors required to perform the active quench and recharge functions.

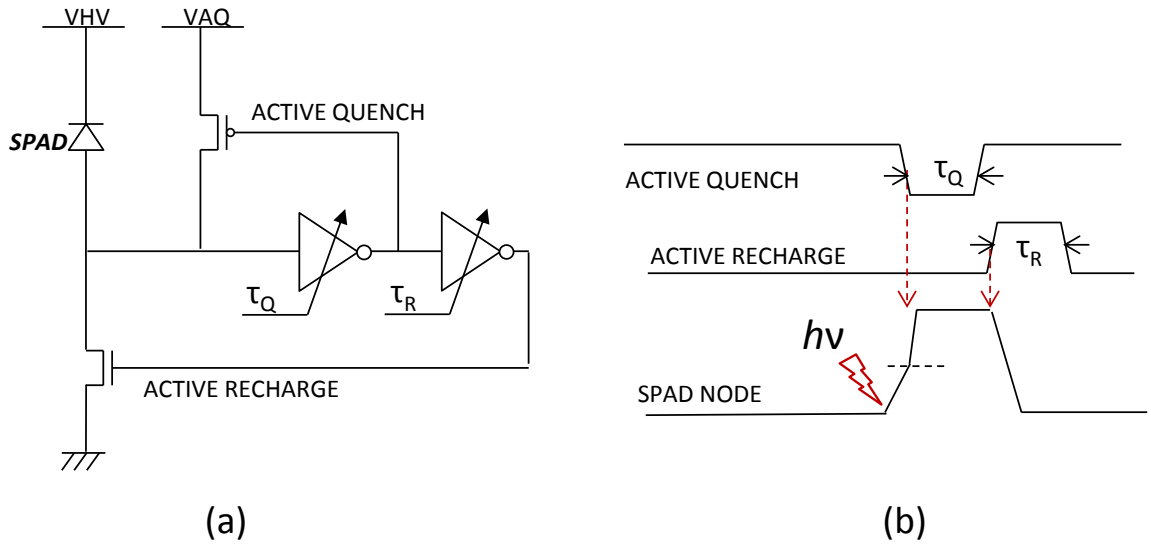


Figure 3.23. Active Quench and Recharge: (a) Generalised circuit schematic showing the delay elements for active quench and recharge. (b) Active quench and recharge timing diagram.

3.3.4. Digital In-Pixel Circuitry

SPADs, when connected through an inverter, become a truly digital imaging device; each photon immediately represented as a digital pulse whose leading edge signals the photon's time of arrival with picosecond precision. It follows that each pixel contains an application specific time conversion or counting circuit that processes every photon incident on the image sensor. Contrary to the prior examples, this class of pixel has in-pixel integration and processing of SPAD events. Being able to count and process each photon in the digital domain mitigates the contribution of transistor-based noise sources from the resulting image (thermal, kT/C, flicker, etc.) leaving only the exposure dependent noise sources: SPAD dark count rate and photon shot noise. Figure 3.24(a) demonstrates a generalised schematic of this pixel category where the digital counter or time conversion logic is not confined to work at the higher excess bias or overvoltage of the SPAD ('VEB'), but may operate at the lowest voltage supply ('VDD') available in the process technology making use of thinner oxide transistors. A generalised layout of the 'digital' pixel is illustrated in Figure 3.24(b), where the photo-sensitive region is marked by the diode symbol highlighting the low fill factor obtainable in such a pixel. In Figure 3.24(c), the same pixel has this restriction lifted in the stacked process.

A number of researchers have explored and made use of this class of pixel. The first was Dr Brian Aull (MIT) demonstrating a 32x32 flash TDC array with hybrid bump bonded stacked SPAD IC [32] in 2002 but first discussed in 1998 in [127]. The first monolithically integrated device was shown by Dr Cristiano Niclass demonstrating the first ITOF SPAD-based pixel in his doctoral research. The 60x48 array was manufactured in 0.35µm technology [36], [121]. The 85µm pixel with 0.5% fill factor has two 8 bit up/down counters synchronised with a sinusoidal LED illuminator.

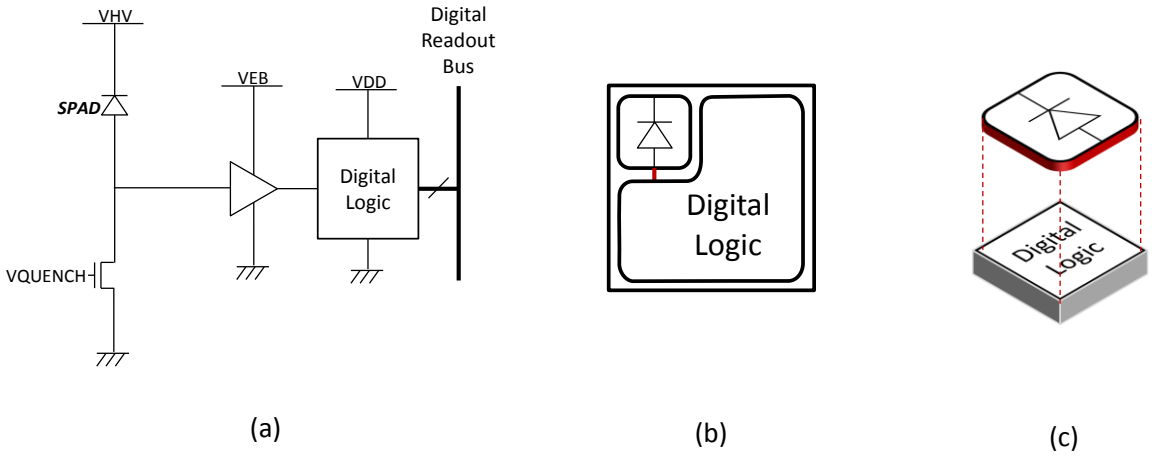


Figure 3.24. SPAD-based pixel with digital logic. (a) Pixel schematic with passive quenched front end buffer and digital logic. (b) Generalised layout in a monolithic process. (c) Generalised layout in a stacked process with digital logic support wafer and photo-sensitive top wafer for the SPAD.

The 'MegaFrame' project created three architectures of digital SPAD pixel for time-resolved FLIM [38], [40], [54], [65]. Two array sizes of these were created: a 32x32 and a 160x128 both in ST's 130nm technology. The 50µm pixel pitch had a 6.6µm active diameter SPAD giving 1.6% fill factor. The three variants were created individually by the three design partners: a 16 element delay line TDC (led by Prof. Edoardo Charbon at EPFL/Tu Delft), a TAC and analogue counter with in-pixel single slope ramp ADC (led by Dr. David Stoppa at FBK) and a gated ring-oscillator (GRO) TDC (led by Prof. Robert Henderson at the University of Edinburgh). As an example, the 32x32 GRO-TDC sensor successfully demonstrated 500k frames per second (FPS) readout of TDC codes. This translates to 488 photons per pixel per second able to be captured and read out. This is suitable for photon-starved imaging, such as FLIM [128]–[130], but as a commercial TOF sensor able to handle ambient light and requiring video-rate output. This sensor type has also been demonstrated in a proof of concept DTOF ranging experiment using a pulsed laser [39].

In his doctoral research [39], Dr Richard Walker from the University of Edinburgh CSS research group, following on from the 'MegaFrame' project, implemented the first in-pixel sigma-delta loop in a 44.65µm digital pixel at 3.1% fill factor for TOF 3D ranging. The 128x96 imager was manufactured in ST's 130nm technology. The 'MiSpia' project produced similar sensors for 3D vision applications. The Milan-based research group have proposed a suite of pixels at 150um pixel pitch in Fraunhofer IMS's 0.35µm technology all with less than 4% fill factor [131]–[134]. Also, a similar work by Rodriguez-Vazquez's research group created a 64x64 array with in-pixel 11b TDC at 64µm at 3.5% fill factor [135].

A comparative table of these works is shown in sections 3.3.7 and 3.3.8. In a monolithic sensor, the area overhead of the in-pixel digital logic significantly impairs and limits the fill factor of these SPAD-based image sensors from reaching more than 10%. This leads to substantial loss of incoming photons from an imaged scene, diminishing the sensitivity of these systems. In a 3D vision system, this is critical to the power consumption, as the illuminator power must be increased to compensate. This low fill factor can be partially mitigated by micro-lensing yet the large pitch still remains, prohibiting high spatial resolution imaging arrays. This class of pixel will suffer from this fundamental problem, with the pitch remaining above 25µm, until chip-stacking technology is available.

3.3.5. In-Pixel Memory

To attain a compact digital pixel, the number of transistors is reduced. The most compact pixel whilst retaining digital functionality is a single bit memory structure. The pixel operates by a SPAD event, within an exposure window, switching the state of the in-pixel memory cell. However, this structure limits the full well of the pixel to one event (a photonic or dark event) and requires an external frame store to build up a digital image. Like the digital-logic class of pixel this removes any systematic noise sources from the pixel and readout; the only noise sources remaining in such a binary image sensor are ideally exposure dependent [61]. Such a SPAD

pixel with a digital single-bit memory offers the first practical step toward realising the digital film sensor or Quanta Image Sensor (QIS) concept proposed by Prof. Eric Fossum in [44], [45], [61] and by extension, as part of this thesis, a Digital Time of Flight Image Sensor. This digital imaging modality is covered in more detail in Chapters 5 and 6 of this thesis, and in [2].

The first example of this time-gated binary pixel architecture was designed by combining a SPAD, an inverter, a time gate and a static memory cell for time-gated FLIM by Maruyama et al. (TU Delft) in a 128x128 array [136][137]. The 25µm pitch in 0.35µm process is a noteworthy reduction compared to the other pixels using digital logic. The 12T NMOS-only pixel uses a NMOS-only SRAM to avoid hot NW spacing rules from the use of PMOS in the pixel circuit. This has the downside of static power consumption during operation. The power consumption of this NMOS-only latch architecture, scales linearly with array size making its use unfeasible for high-resolution low power image sensors. Recently this architecture was scaled by the same group, published by Burri et al., to a 512x128 array and demonstrated successfully for FLIM in [138] and more recently in [139]. The use of dynamic memory or CMOS static memory removes the static power consumption and lifts the scalability limit from those works. A CMOS single bit latch pixel was implemented by Aull et al. in their 256x256 array in [140]. Alternatively an NMOS-only dynamic memory circuit may be used at the cost of limited memory retention time (in the order of milliseconds). Figure 3.25(a) shows a simplified dynamic memory circuit diagram alongside a generalised layout view in Figure 3.25(b), the memory logic occupies a considerably reduced area to the digital logic of the previous examples of digital pixels. In a stacked process, the lower logic layer may act as a frame store to digitally oversample each memory pixel. An 'N' by 'N' set of binary memory pixels may be summed together in a digital oversampling block creating a macro-pixel. Figure 3.25(c) illustrates a macro-pixel consisting of 4x4 memory pixels in the top layer, and single digital oversampling block in the bottom logic layer.

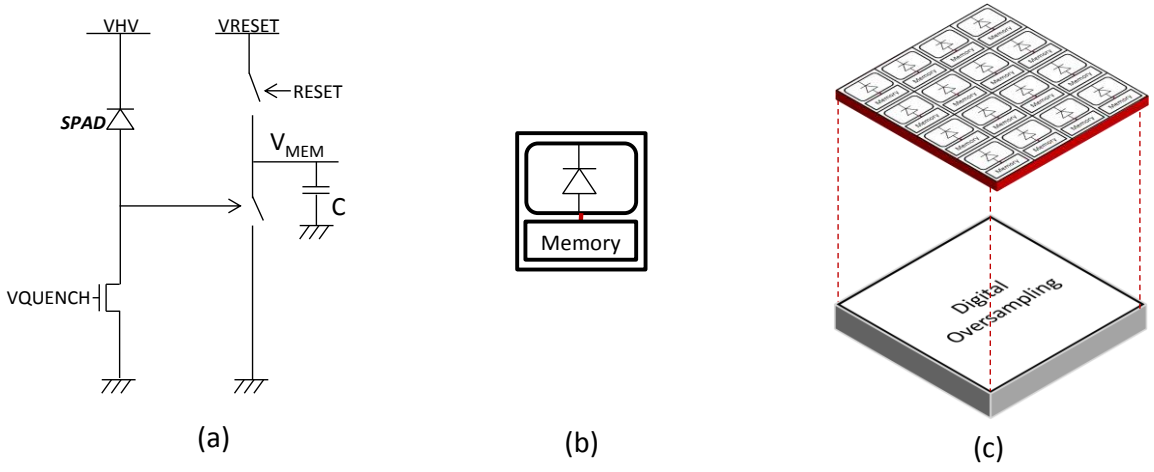


Figure 3.25. In-pixel memory: (a) Simplified dynamic memory circuit. (b) Pixel layout illustrating reduced digital logic area. (c) One of the many possibilities for a single bit memory pixel grouping in a stacked process with digital oversampling logic under a 4x4 sub-pixel array or 'jot' array [45]**.**

3.3.6. Analogue In-Pixel Circuitry

Research within this thesis and recent research by other groups has investigated two analogue pixel approaches for high resolution SPAD-based image sensors: TAC pixels for TCSPC imagers, and analogue counters for time-gated single photon counting imagers. Considering the latter, Figure 3.26 illustrates the simplified circuit diagrams of the two different methods of analogue counting used in the counting pixels described in this section. Figure 3.26(a) shows the switched current source (SCS) and Figure 3.26(b) displays the charge transfer amplifier (CTA) circuit. An analogue counter operates by a SPAD pulse crossing a threshold, switching the analogue counter to activate its counting function. As described in the theoretical equation in Figure 3.26(a), the incremental voltage step of the switched current source counter (ΔV_C) is proportional to the current through the current sink and the SPAD recharge time. On the other hand, the voltage step of the CTA circuit is proportional to the capacitance ratio and independent of the SPAD recharge time. The full operation and analyses of these circuits are described in greater detail in Chapter 4.

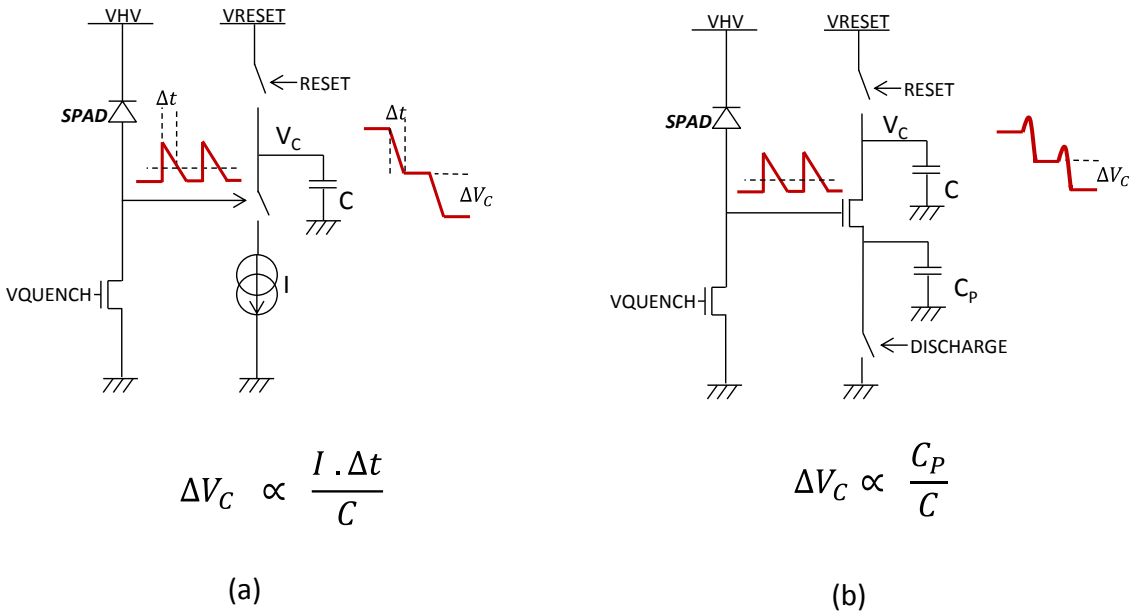


Figure 3.26. Simplified circuit diagrams and theoretical equations for the operation of two analogue counter implementations: (a) the switched current source and (b) the charge transfer circuit.

Stoppa et al. (FBK) were the first to report an analogue counter based ITOF approach for a time gated SPAD pixel in 2007 [66]. The pixel was implemented in a line sensor format of 64x1 pixels, at a pixel pitch of 38x180µm in a 0.8µm high voltage process. With 25 CMOS transistors this sensor provided a proof of concept that a SPAD-based ITOF pixel could be implemented with an analogue counter using a low number of transistors. The pixel circuit was initially designed as an integrating TAC but then re-purposed to be a SCS counter with a maximum counting capability of 10,000 SPAD avalanches or events. The discharge time ('Δt') is controlled by the dead time of the SPAD. The SPAD dead time is proportional to the SPAD excess bias and

the quench voltage, with a transistor based quench making the discharge time variable. This links the variability of the counter step with the SPAD dead time variability which is undesirable.

That team expanded this work by implementing the analogue variant of the 'MegaFrame' pixel, as previously mentioned [141]. The pixel had both a similar combined TAC and analogue counter core and a single slope ramp 8b ADC with two 8b memories all contained in the 50µm pixel. This pixel concept has the benefit of being multi-functional but suffers from the same low fill factor issue as the digital class of pixel. Again a switched current source was employed, with a maximum of 40 counts within a 2µs frame time (500kFPS). Contrary to their previous work the discharge time is controlled by a digital logic block and is independent of SPAD dead time, with the intention of ensuring high uniformity in the count step sizes across the array. However, measured variability results were not presented in their work.

Chitnis et al. (Oxford) proposed a 10T counter pixel for a SPAD-based image sensor in [142]. The pixel was manufactured in a 0.18µm UMC standard CMOS process having a 10µm diameter active area SPAD, 30µm pitch with 8.7% fill factor [143]. Although the use of three PMOS devices in the pixel limits achieving a higher fill factor, this is the first example of a square image sensor pixel with a low number of transistors. They implement two distinct counter operations dependent on external bias voltage input: a linear single slope mode and a two-slope logarithmic mode. The step size achieved was ~140mV per SPAD event in the linear mode with a constant reference voltage. An approximately logarithmic counter response was achieved by modulating a MOS capacitance proportional to the counter voltage. Furthermore, like the first example, the SPAD dead time sets the discharge period, adding a source of variability into the counter operation.

Pancheri et al. (FBK) improved their previous two works trialling a 32x32 array of 25µm pitch analogue single photon counting pixels at an impressive 20.8% fill factor in a 0.35µm high voltage process [57][144]. This first test array had a nanosecond time gating delay generator on the chip facilitating taking FLIM images by means of multiple exposures with a progressively delayed time gate. This 12T pixel circuit was the first example of a pixel employing only NMOS transistors to achieve a high fill factor. They employ a switched current source counter with a front end gating circuit to generate a picosecond duration input pulse removing the dependence on SPAD dead time. An NMOS-only inverter is used in the gating circuit with a static power consumption making it unsuitable for scaling to larger arrays. High 11% photo-response non-uniformity (PRNU) is recorded across the tested array, with typical 10mV voltage step although the source of the non-uniformity is not discussed in the work.

In all of the previous examples, the discharge current is set through either a voltage bias or a current mirror but none presents voltage step tunability. Panina and Pancheri et al. [145] (FBK/Trento) describe simulation results of two pixel designs each with tuneable voltage step capability, the first is a SCS similar to the previous work and the second is an initial example of a CTA proposed for use in an analogue SPAD pixel. The voltage step range of the SCS is 5mV to 20mV per SPAD event but the variable CTA step range is not simulated. In

the two designs, the authors propose replacing the NMOS-only inverter with a CMOS inverter and adding two more PMOS devices without considering the effect on the pixel fill factor. In their work, the central conclusion is made by presenting side by side evaluations of the simulated voltage step variability of the two different approaches. The standard deviation of the SCS is in the order of 7 to 9 percent whilst the CTA is less than 1.2%.

Panina and Pancheri et al. presented the measured results of these counter pixels at SPIE Optics and ESSCIRC in 2013 [146], [147]. A set of 40 pixel linear test arrays were manufactured in a 0.15µm process (LFoundry). Although the pitch was not published, it has been estimated by the author at 40x20µm with 9.8% fill factor. The CTA and SCS counters described in [145] were tested in [146] and [147] respectively with a SPAD input demonstrating 1.9% CTA voltage step variability or PRNU and 8.6% SCS PRNU which is in agreement with their simulated work. A modified version of the CTA was published in [148], removing the time gate and replacing it with two chained CMOS inverters. It was tested with a digital pulse generator and is shown to have consistently less than 4% voltage step variability across 4 to 14mV per step range.

An important note in terms of project timeline must be made here for clarity. The work by Panina and Pancheri et al. [145] was published in June 2012 after our initial research and design of a CTA-based SPAD pixel with bias controlled voltage step sensitivity which was started in September 2011 and completed by March 2012. The CTA-based pixel test array was manufactured in April 2012, but the measured results were published in June 2013 at the International Image Sensors Workshop [1]. The conclusion that the CTA is less variable than the SCS is in agreement with the findings of the FBK team's research that arrived at a similar conclusion made in Chapter 4 and [1], [2]. Recently in 2015, the same group has published a CTA based imager at 160x120 resolution in [149].

The reduction in pixel pitch that an analogue approach offers, was seized upon in the course of this thesis research work and in the parallel research into SPAD-based TAC pixels by colleague Luca Parmesan. In this thesis work, a 9.8µm pitch CTA analogue counter pixel is described in Chapter 4 and in [1] with 3.1% fill factor and bias controlled voltage step. At the time of publication, this was the smallest pitch SPAD-based image sensor pixel. In collaborative doctoral research with Luca Parmesan, a novel architecture of sample and hold TAC pixel was demonstrated in a similar 9.8µm pitch using the same SPAD with 3.1% fill factor in [4]. A second revision of the CTA counter pixel was resized to 8µm pitch and increased to 26.8% fill factor, and implemented in a QVGA array which is described in Chapters 5, 6 and [2]. Again at time of publication, it was the smallest SPAD imager pixel design achieved by using the double-strip local well sharing technique. A comparison of SPAD-based analogue counting pixels is presented in Table 4.8 in Chapter 4. This pixel layout method was used, again in collaborative doctoral research with Luca Parmesan, to create a 256x256 sample and hold TAC-based imager for FLIM with 20.8% fill factor at 8µm pitch. In comparison to the all-digital TCSPC pixels, this pitch reduction and fill factor improvement by using analogue in-pixel circuitry facilitated the first high-resolution 256x256 TCSPC image sensor in [7].

3.3.7. Pixel Pitch and Fill Factor Comparison

Comparing both SPAD-based DSiPM and imager pixels, Table 3.3 lists the technology node, pixel pitch and fill factor of the referenced works in the six categories of SPAD pixel architectures. The array architecture of the works is listed alongside and the image sensor architectures are highlighted in grey. These data are graphed in Figure 3.27 displaying the pixel pitch against fill factor with the pixel categories highlighted. The pitch data are calculated as the square root of the area to equalise all pixels into a square format. Figure 3.28 shows the same examples graphed again with the pitch (square root of area) normalised to a nominal 0.1µm process to account for the different process technology nodes to offer some compensation for transistor sizing. Some of the references from the table do not give complete information and so are not included in the graphs.

The 'SPAD only' devices maintain the highest pixel fill factor with no area overhead of in-pixel logic showing the limits of what is attainable using global well sharing pushing the SPADs as close together as possible. This indicates what will be possible using chip stacking technology. The 'minimal logic' pixels illustrate the small pitch that is attainable with a few transistors. There is a clear distinction of the 'Digital' and 'Active Quench' pixels from the other categories with large pitch and low fill factor (with the exception of the Phillips/NXP PET Sensor) due to the inclusion of digital logic in-pixel. On the other hand, the pared down 'in-pixel memory' single bit time-gated digital pixels reveal the pitch attainable by optimising the pixel design.

To achieve a high resolution time-resolved SPAD-based image sensor architecture capable of being scaled to megapixel arrays, photon counting must be performed in-pixel (and not row by row at the edge of the array) ruling out both the 'SPAD only' and 'Minimal' categories. The large pitch of 'Digital' and 'Active Quench' pixels eliminates them from being successfully applied in a high resolution array in a monolithic process. An analogue pixel offers in-pixel processing of SPAD events with a low number of transistors whilst occupying a small area and reaching high fill factor. The analogue approach provides an optimal solution for implementing such an image sensor and is chosen to be built upon to form the basis of this thesis research.

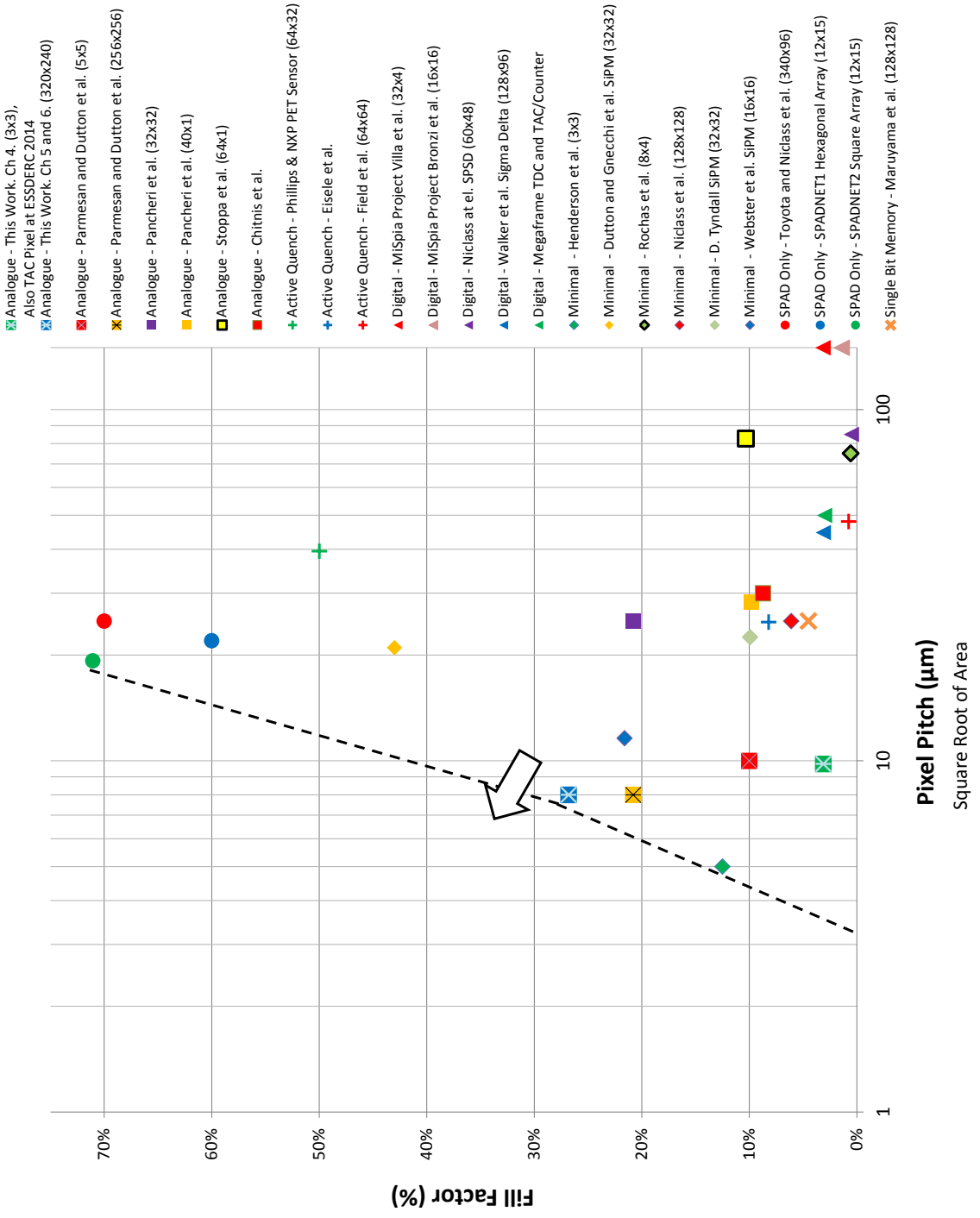


Figure 3.27. Graph of pixel pitch (on a log scale) against fill factor for a range of SPAD-based sensors with a line showing the boundary of the best case pitch versus fill factor and an arrow showing the future direction of higher fill factor and lower pitch pixels in SPAD-based image sensors.

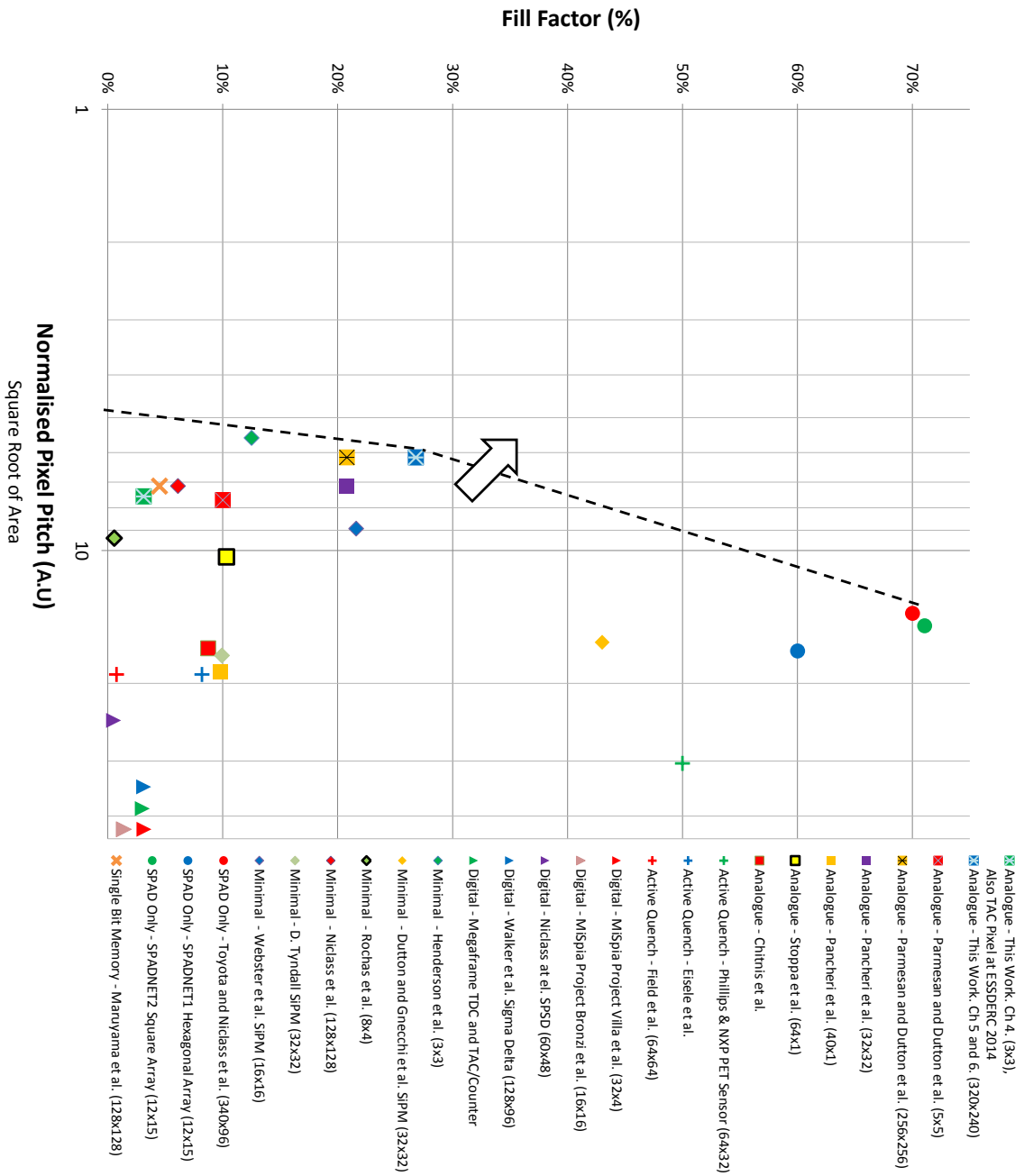


Figure 3.28. Graph of normalised pixel pitch (on a log scale) against fill factor for a range of SPAD-based sensors with a line showing the boundary of the best case pitch versus fill factor and an arrow showing the future direction of higher fill factor and lower pitch pixels in SPAD-based image sensors.

Ref	Type	Title	Year (20XX)	CMOS Process Node (µm)	Pitch X (µm)	Pitch Y (µm)	Sq. Root of Area (µm)	Normalised Pitch* (AU)	Fill Factor (%)	Pixel Architecture
[1]	Analogue	This Work. Ch 4. SPC (3x3)	13	0.13	9.8	9.8	9.8	7.5	3.1	Img
[2]	Analogue	This Work. Ch 5 and 6. SPC (320x240)	14	0.13	8.0	8.0	8.0	6.2	26.8	Img
[4]	Analogue	Parmesan and Dutton et al. S/H TAC (3x3)	14	0.13	9.8	9.8	9.8	7.5	3.1	Img
[150]	Analogue	Parmesan, Integrating TAC and SPC (5x5) *(Unpublished)*	15	0.13	10.0	10.0	10.0	7.7	10.0	Img
[7]	Analogue	Parmesan and Dutton et al. S/H TAC (256x256)	15	0.13	8.0	8.0	8.0	6.2	20.8	Img
[66]	Analogue	Stoppa et al. SPC (64x1)	07	0.8	38.0	180	82.7	10.3	10.3	Line
[57]	Analogue	Pancheri et al. SPC (32x32)	11	0.35	25.0	25.0	25.0	7.1	20.8	Img
[146]	Analogue	Pancheri et al. SPC (40x1)	13	0.15	40.0	20.0	28.3	18.9	9.8	Img
[143]	Analogue	Chitnis et al. SPC	10	0.18	30.0	30.0	30.0	16.7	8.7	Img
[83]	Active Quench	Phillips & NXP PET Sensor (64x32)	09	0.13	52.0	30.0	39.5	30.4	50.0	SiPM
[89]	Active Quench	Eisele et al.	11	0.13	22.0	28.0	24.8	19.1	8.2	SiPM
[125]	Active Quench	Field et al. (64x64)	13	0.13	48.0	48.0	48.0	36.9	0.8	Line
[103]	Active Quench	Finkelstein et al.	07	0.18	180	96.0	131.5	73.0	0.5	SiPM
[131]	Digital	MiSpia Project Villa et al. (32x4)	12	0.35	150	150	150.0	42.9	3.1	Img
[132]	Digital	MiSpia Project Bronzi et al. (16x16)	12	0.35	150	150	150.0	42.9	1.4	Img
[121]	Digital	Niclass at el. SPSD (60x48)	05	0.35	85.0	85.0	85.0	24.3	0.5	Img
[53]	Digital	Walker et al. Sigma Delta (128x96)	11	0.13	44.7	44.7	44.7	34.3	3.1	Img
[38]	Digital	Megaframe TDC and TAC/Counter	09	0.13	50.0	50.0	50.0	38.5	3.0	Img
[72]	Minimal	Rochas et al. (8x4)	03	0.8	75.0	75.0	75.0	9.4	0.6	SiPM
[151]	Minimal	Henderson et al. (3x3)	09	0.09	5.0	5.0	5.0	5.6	12.5	SiPM
[5]	Minimal	Dutton and Gnecchi et al. SiPM (32x32)	14	0.13	21.0	21.0	21.0	16.2	43.0	SiPM
[56]	Minimal	Niclass et al. (128x128)	08	0.35	25.0	25.0	25.0	7.1	6.1	ED
[119]	Minimal	D. Tyndall SiPM (32x32)	12	0.13	22.5	22.5	22.5	17.3	9.9	SiPM
[95]	Minimal	Webster et al. SiPM (16x16)	12	0.13	11.6	11.6	11.6	8.9	21.6	SiPM
[35]	SPAD Only	Toyota and Niclass et al. (340x96)	13	0.18	25.0	25.0	25.0	13.9	70.0	Line/ ED
[41]	SPAD Only	SPADNET1 Hexagonal Array (12x15)	12	0.13	19.3	19.3	22.0	16.9	60.0	SiPM
[152]	SPAD Only	SPADNET2 Square Array (16x16)	14	0.13	19.3	19.3	19.3	14.8	71.1	SiPM
[117]	SPAD Only	Hamamatsu MPPC (300x300)	13	-	10.0	10.0	10.0	-	-	SiPM
[136]	Single Bit Memory	Maruyama et al. (128x128)	11	0.35	25	25	7.14	4.5	5	Img

Table 3.4 SPAD Image Sensors Table listing the Pixel Category, Pixel Pitch, Fill Factor and CMOS Technology Node. *Square Root of Pixel Area Normalised to a 0.1µm CMOS Process. ED = Event Driven, Img = Image Sensor, Line = Line Sensor, SiPM = DSiPM or ASiPM

3.3.8. Further Pixel Comparison

Further to pixel pitch and fill factor comparison, the following two tables seek to highlight the differences between DTOF digital pixel based image sensors in Table 3.5, and time-gated single photon counting pixels in Table 3.6. The following DTOF sensors require a large frame store (either on-chip or on FPGA) for per-pixel histogram generation, whereas the ITOF sensors described in Table 3.5 have in-pixel counting or averaging which is ideal for a time-resolved imaging sensor as it massively reduces the output data rate.

The analogue pixels exhibit a compromise of low pitch and high fill factor whilst still maintaining a degree of in-pixel function. To make a comparison in the same process node, Pancheri's 25µm analogue counter pixel achieves similar performance with 8bit (256 step counter) at 97% less area with 19% higher fill factor than Bronzi's 150µm digital pixel from the MiSpia project with two 8 bit counters.

First Author Surname	Niclass	Niclass	Villa	Richardson	Veerappan
Institution	EPFL	Toyota	Polimi	U Edinburgh /ST	TUDelft
Reference	[56]	[35]	[131], [134]	[106]	[54]
Year Published	2008	2013	2012	2010	2011
Process Node	0.35µm	180nm	0.35µm	130nm	
Array Size	128x128	340x96	32x4	32x32	160x128
Readout Architecture	Event Driven Image Sensor	Line	Image Sensor	Image Sensor	
Pixel Type	Minimal	SPAD Only	Digital	Digital	
Time Conversion Circuit	Flash TDC	Flash TDC	Flash TDC	GRO TDC	
In Pixel or Column	1 per 4 Columns	Column	In-Pixel	In-Pixel	
Pixel Pitch (µm)	25	25	150	50	
Fill Factor (%)	6.1	70	3.14	1	

Table 3.5. Comparison Table of SPAD-based DTOF sensors.

First Author Surname	Niclass	Stoppa	Walker	Bronzi	Pancheri
Institution	EPFL	FBK	UoE / ST	Polimi	FBK
Reference	[55]	[66]	[39]	[132]	[144]
Year Published	2006	2007	2011	2012	2013
Type of TOF	Indirect	Indirect	Indirect	Indirect	Indirect
Process Node	0.35µm	0.8µm	130nm	0.35µm	0.35µm
Array Size	60x48	64x1	128x96	16x16	32x32
Pixel Type	Digital	Analogue	Digital	Digital	Analogue
Time Conversion Circuit	8 bit Counters	Analogue Counter	6 bit Sigma Delta	Counters	Analogue Counter
In Pixel or Column	In-Pixel	In-Pixel	In-Pixel†	In-Pixel	In Pixel
Pixel Pitch (µm)	85	38 x 180	44.65	150	25
Fill Factor (%)	0.5	10.3	3.1	1.39	20.8

Table 3.6. Comparison Table of SPAD-based ITOF sensors. † With Column Level 10 bit Decimation

3.4. Summary and Discussion

In this chapter, the wide range of CMOS SPAD diode and pixel circuit implementations were compared, in order to work toward design the optimal SPAD-based time-gated single photon counting image sensor pixel.

Quench and recharge circuits and SPAD enabling and disabling logic were reviewed. SPADs with a transistor based quench are favoured as those SPADs with a PQPR resistor cannot be disabled or incorporated into active quench and recharge functionality. The single NMOS PQPR transistor is preferred, for a scalable image sensor pixel, as PMOS devices increase the pixel pitch due to NW spacing rules. Also, active circuits require many CMOS transistors and the stability of a large array of AQAR circuits operating at high frequency simultaneously has yet to be studied and requires further investigation.

Directly connected SPADs are advantageous over capacitively coupled SPADs, as the latter have the drawback of charge loss, and voltage pulse attenuation, due to the parasitic capacitor divider requiring the SPAD to be operated at a higher excess bias than a directly connected device to obtain the same voltage swing to trigger the SPAD pixel logic. However, the attenuation effect may be reduced using the proposed MOM coupling capacitor, yet the pixel and TOF system performance may still be impeded by capacitor variability, optical interference and increased SPAD dead time.

Global and local well sharing techniques can be applied to substrate isolated SPAD structures to improve the fill factor and lower the pixel pitch. Global well sharing is ideal for stacked technology, but in a monolithic process is restricted to DSiPMs and low resolution imaging arrays. However, for a scalable monolithic pixel, the double strip local well sharing approach provides the greatest fill factor at the expense of different horizontal and vertical MTFs.

Substrate isolated and non-substrate isolated diode structures were modelled for imaging at 850nm. The double-strip substrate isolated SPAD pixel has higher PDE over the non-substrate isolated 'deep' SPAD version at pixel pitches below 13µm.

The positive drive substrate isolated Richardson PW to DNW SPAD structure is selected for an image sensor pixel due to a number of factors: the amenability to well-sharing, low DCR, direct anode to pixel circuit connection, and the highest PDE at 850nm of comparable substrate isolated structures.

In terms of pixel circuits, the many different CMOS SPAD pixel circuit designs were considered. The voltage output of the SPAD is inherently a spiking digital pulse and the ability to process every photon in the digital domain offers the ideal condition for shot-noise limited single photon counting imaging. However, monolithic all-digital pixels suffer from large pitch leading to low spatial resolution arrays with low fill factor and limited sensitivity. In a monolithic technology an analogue pixel offers the most compact and high fill factor solution providing both time-gating and photon counting capability. All-digital pixels will be able to compete in pitch and fill factor as chip stacking technology becomes available. Operating in the analogue domain introduces

analogue noise sources (kT/C, thermal, etc.) so care must be taken in the pixel and readout design to minimise the impact of these. A further advantage of using an analogue pixel, is the compatibility with existing CIS readout architectures with column parallel ADCs and corrective image signal processing (ISP).

Although a SPAD-based pixel structure will not shrink to sub-micron pixel pitch, an array of SPAD-based single bit digital pixels offers a look ahead to the properties of Fossum's proposed oversampled binary image sensor, the Quanta Image Sensor (QIS).

4. An Analogue Counter CMOS SPAD Pixel

4.1. Introduction

This chapter describes the design, implementation and characterisation of a three by three pixel array envisaged to prove the concept that a time-gated SPAD-based analogue single photon counting pixel could be designed to be scalable, in pixel pitch and current consumption, in order to realise large imaging arrays. This test array had the following primary goals:

- To investigate the performance of an analogue single photon counting (SPC) pixel circuit with input from either SPAD or FPGA.
- Decrease the pixel pitch from a state of the art 25μm to below 10μm to be competitive with other imaging technologies.

Further to these aims, TAC pixel operation and the choice of diode were also investigated in parallel experiments described in Appendix 1. A number of TAC pixels were implemented alongside the primary SPC pixels. These were conceived by Prof. Robert Henderson to create a time correlated single photon counting (TCSPC) image sensor, and this topic became the doctoral research of Luca Parmesan. These TAC pixels are detailed in Appendix 1 and the characterisation of these TCSPC pixels was published in [4]. Further to the pixel circuit design, two SPAD constructions were trialled: the Richardson PW to DNW 'Shallow' SPAD and the Webster DNW to Substrate 'deep' SPAD. However, the front end NMOS-only time gate structure used on all the 'deep' SPAD pixels encountered the same design fault which prohibited characterisation of these experiments. The 'deep' SPAD pixel circuits and the design fault are also described in Appendix 1.

4.2. Counter Design and Simulation

This section details the design and simulation of the eleven transistor time-gated analogue counter circuit. There are two parts to the circuit: the first is the front-end stage, consisting of quench, time gate and test input; the second is the analogue integrator and APS-style readout. The central limitation on the circuit design is the confined physical area available for transistors. Any transistor which may be sized to minimum dimensions (without severe impairment to pixel variability) is restricted in area, to allow the counter structure to use all the remaining physical space to achieve the highest possible performance in terms of both counter depth (dynamic range) and minimisation of pixel to pixel photo response non-uniformity (PRNU).

4.2.1. Quench, Time Gate and Test Input

Figure 4.1 illustrates the design of the front-end circuit consisting of the SPAD, single transistor PQPR quench and disable pull-up. The two-transistor time gate, consisting of an 'enable' inline switch and a 'disable' pull-down, is a novel structure. The test-input pull-up facilitates disabling the time-gated SPAD input and permits an electrical test pulse to trigger the counter. In this test operation, the time-gate 'enable' is held low and the time-gate 'disable' pull-down is fed with the opposite polarity of the same trigger signal. The supply name 'VRT', used in the figure, is an abbreviation of 'V Reset' which is the common supply name in conventional CIS pixels. Table 4.1 describes the front end device sizes. With the exception of the quench transistor M1, all four other devices are minimum sized. This minimum width is critical as these devices are placed in a layout strip between SPADs and any increase in width would increase pixel pitch. However, the PQPR transistor length was maximised to approximately 4µm to occupy the remaining available area and to minimise resistance variability.

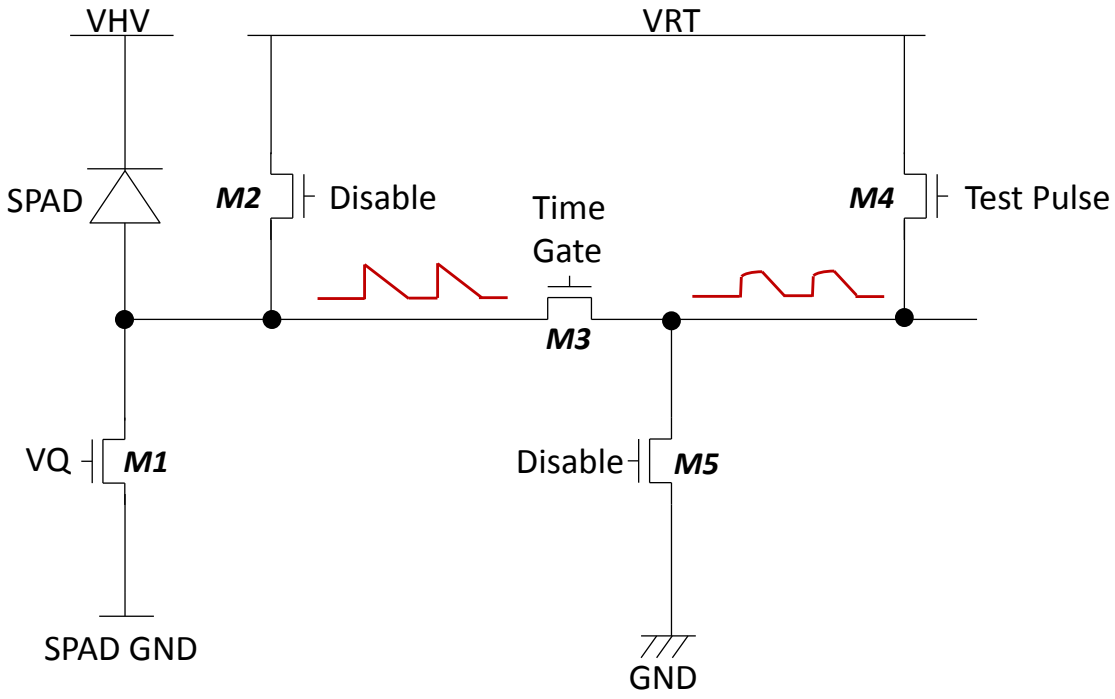


Figure 4.1. Front-end SPAD quench, time-gate and test input circuit.

Transistor	Description	Width (µm)	Length (µm)
M1	SPAD PQPR Transistor	0.3	3.94
M2	Time Gate Enable Inline Switch	0.3	0.35
M3	Time Gate Disable Pull Down	0.3	0.35
M4	SPAD-Node Pull Up	0.3	0.35
M5	External Test Input	0.3	0.35

Table 4.1. SPAD pixel front-end transistor sizes.

Figure 4.2 illustrates the first of two transient simulations of the two transistor time gate. The time gate enable and disable signals are generated through complementary 3.3V logic. The SPAD is simulated with a 1V peak triangular pulse voltage source. The first five SPAD pulses in the simulation are isolated from the counter. At the edge of the time gate, the disadvantage of this design is evident as the sixth pulse is clipped and the trailing SPAD recharge is captured. This is an unavoidable consequence of this design and will temporally blur the leading edge of the time gate. The next six pulses successfully pass through to the counter input. The final pulse is again temporally sliced in the falling SPAD recharge. The second transient simulation in Figure 4.3 has a 3V excess bias SPAD pulse, which is limited, by the switch maximum voltage of $V_{TimeGate}$ (3.3V) minus the threshold V_{TM2} of the switch (~0.6V), to 2.4V at the counter input.

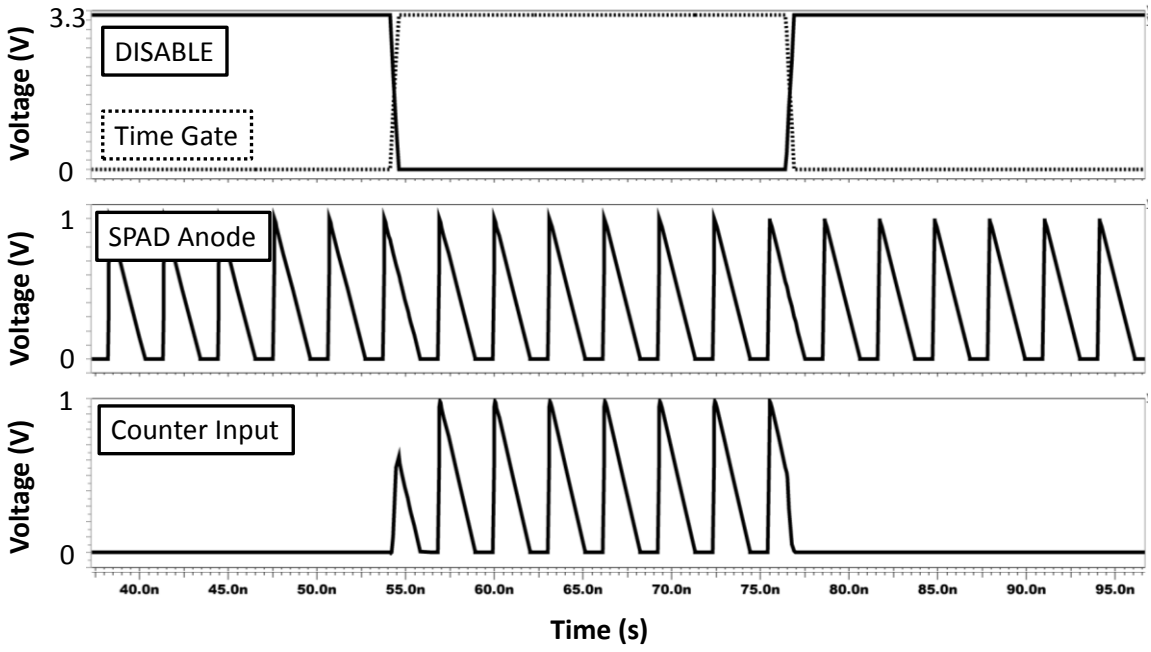


Figure 4.2. Transient simulation of 2T time-gate. SPAD anode 1V excess bias. Time gate 21.8ns gate width and 500ps rise and fall times with 3.3V input.

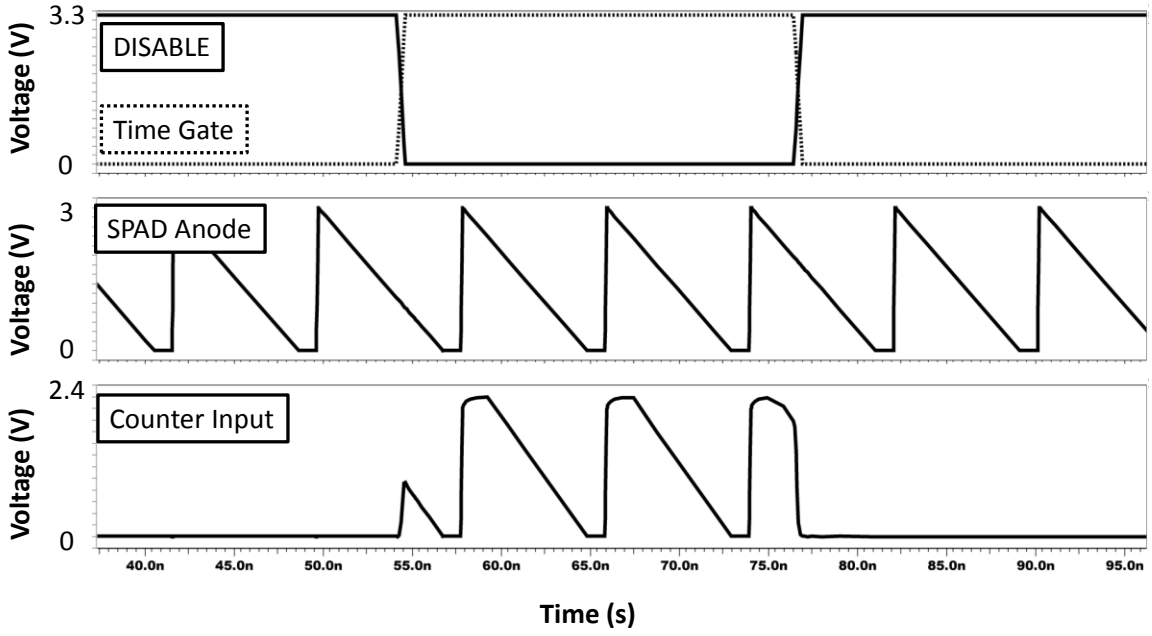


Figure 4.3. Transient simulation of 2T time-gate. SPAD anode 3V excess bias. Time gate 21.8ns gate width and 500ps rise and fall times with 3.3V input.

4.2.2. Analogue Single Photon Counter

Section 3.2.3 described the two different architectures of analogue counter SPAD pixels: the switched current source and the charge transfer amplifier. There are several downsides to implementing one of the existing SCS or CTA circuits. The SCS implementations [57], [142], [143] suffer from high variability which could be compensated per pixel with an auto-zero switched capacitor circuit per pixel. On the other hand, Pancheri's CTA requires two complementary logic signals ('*CHARGE*' and '*DISCHARGE*') from two inverters [146]. All these prior counter circuits require PMOS logic to implement, lowering pixel fill factor. Instead, this is addressed in this work by a simple cross between the two circuits; a NMOS only pixel design is proposed which is a hybrid of the two architectures achieving bias controlled voltage step, an adjustable full well (maximum counts) and low variability. Figure 4.4 describes the genesis of the hybrid counter architecture. The design combines a SCS with bias controllable voltage step (Figure 4.4(a)) and a CTA (Figure 4.4(b)), creating a hybrid SCS and CTA analogue counter shown in Figure 4.4(c). The dominant mode of operation is controlled by the two bias voltages V_G and V_S.

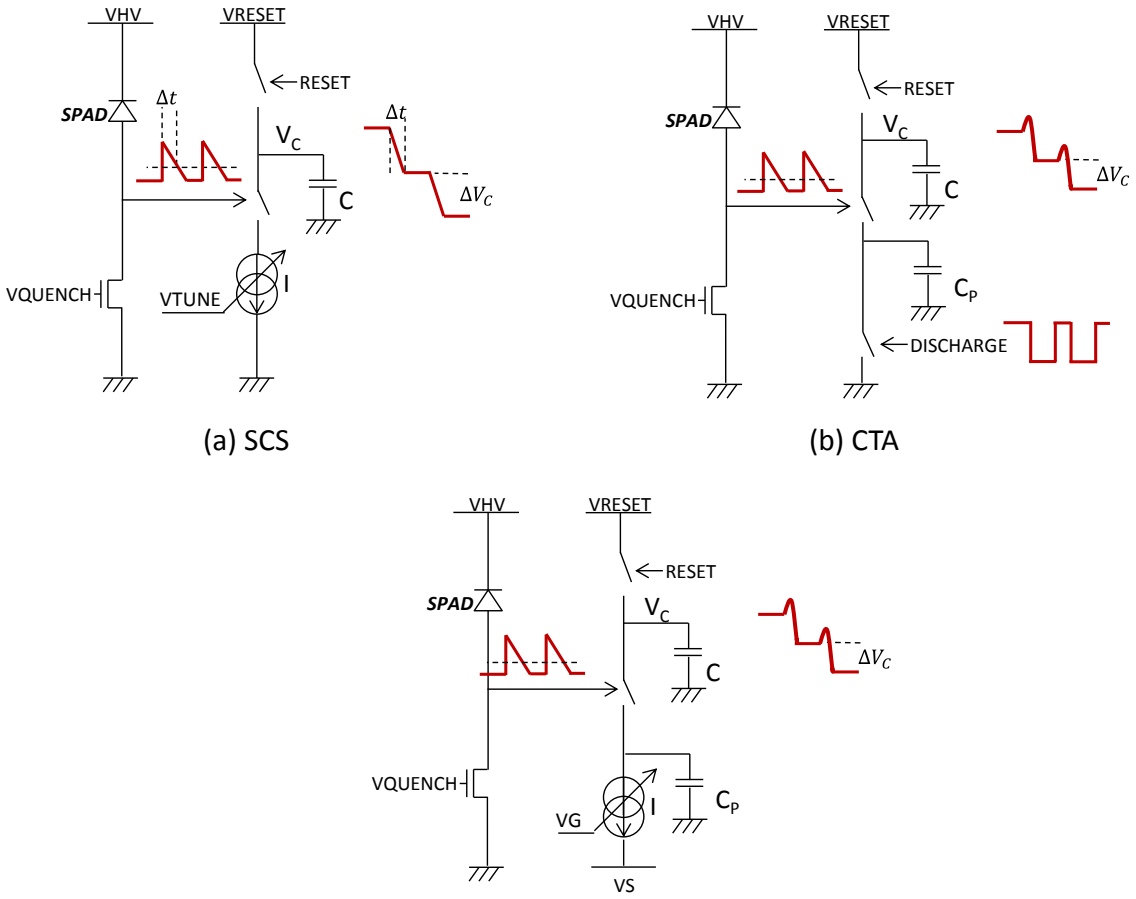


Figure 4.4. Three diagrams showing the genesis of the analogue counter design. (a) Bias controllable SCS, (b) CTA, and (c) the hybrid proposed approach.

4.2.2.1. Proposed Hybrid Analogue Counter Circuit

The hybrid counter pixel proposed is shown in Figure 4.5. The analogue integrator structure consists of M7 dynamic source follower, discharge transistor M8 and poly capacitor MC. The counter design integrates into the conventional NMOS CIS APS readout of M9 source follower, M10 read select and M6 reset, the classic 3T structure proposed by Mendis in [153] following Noble's 1968 array readout paper [28]. The SPC operation is intended to function in either CTA mode, with small voltages steps (µV to mV range) and a resulting large counting capability or full well, or SCS mode with large voltage steps (100's mV to V's) and a small full well.

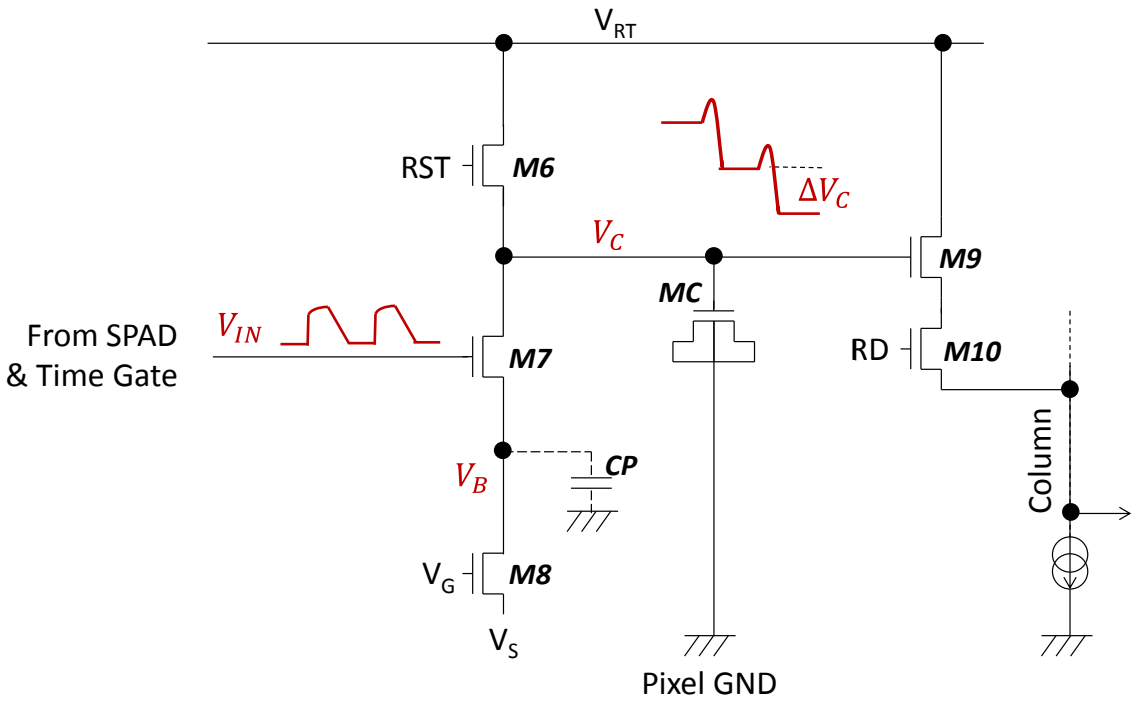


Figure 4.5. Proposed hybrid analogue counter pixel with conventional 3T APS readout.

The CTA was originally proposed by Kotani et al. as an ultra-low power amplifier architecture for a CIS column parallel comparator [59]. Marble et al. revisited the circuit for ultra-low power ADCs in [154], [155] and jointly with Kotani in [156]. In this work, only the NMOS CTA is used and the active switching logic removed to create a novel passive CTA architecture.

Initially in both modes, the pixel is reset by asserting input RST high, charging the capacitance C_T (the capacitance of MC above threshold) to the level of the supply V_{RT}. The maximum voltage, the reset switch can pull the capacitor to, is V_{RST} minus the threshold voltage V_{TM6}. For this inline switch and capacitor arrangement, setting V_{RST} 300mV over the process maximum to 3.6V, sets the maximum capacitor voltage to 2.7V which correspondingly sets the V_{RT} supply voltage to 2.7V. At the beginning of the reset pulse, M6 is in saturation for a short duration before moving into sub-threshold for the majority of the reset period. Once the capacitor is fully charged, RST is asserted low, permitting the pixel to begin SPC operation.

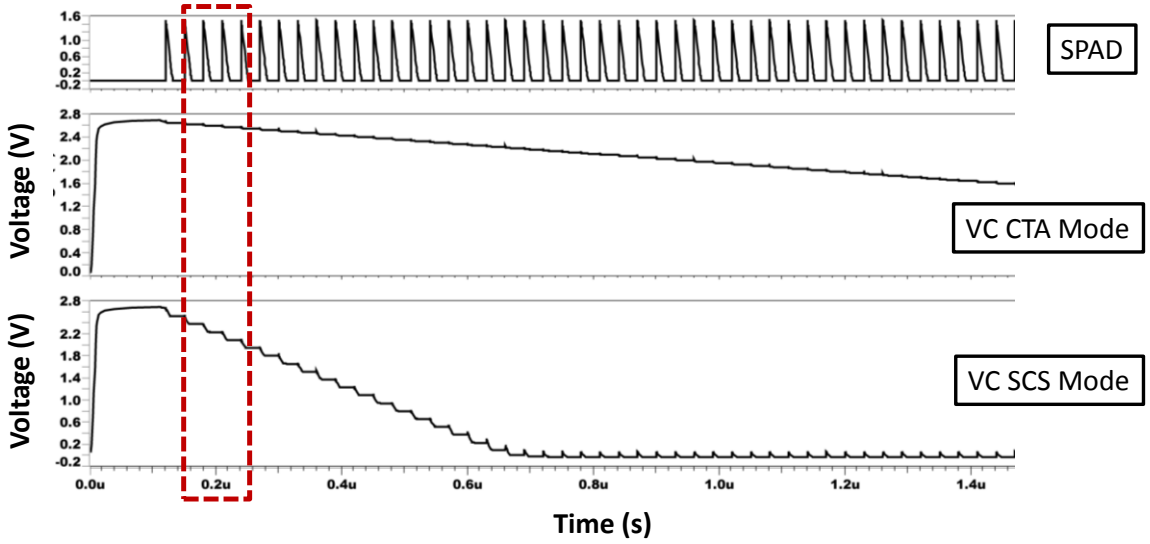


Figure 4.6. Transient simulation of the hybrid analogue counter showing CTA mode and SCS mode. The red highlight box indicates the zoom in region of the next two figures.

Figure 4.6 illustrates a transient simulation of the hybrid analogue counter pixel in both CTA and SCS modes. The top waveform shows a repetitive SPAD pulse with 10ns dead time and idealised 20ns repetition period. The second waveform shows CTA operation with 60 steps at V_G = 0.43V, V_S = 0V with M8 biased sub threshold in weak inversion. The nodes RST, V_{IN}, and V_B are included in Figure 4.5 alongside V_C. The third waveform shows SCS operation with 20 steps at V_G = 0.6V, V_S = 0V with M8 biased above threshold in strong inversion.

Figures 4.7 and 4.8 show a zoomed in section of the transient simulation, first showing the voltage on node V_C, then secondly showing the current through M7 and the voltage on node V_B in the two counter modes. In SCS mode, V_B tracks the SPAD node showing source follower operation. After the initial charge spike on V_B, the I_{DSM7} current is constant for the period where the SPAD counter input is above the M7 threshold. In CTA mode, node V_B is charged up and as seen from the current in I_{DSM7}, M7 is in the cut-off region. The discharge of node V_B is longer and not equal to the dead time of the SPAD.

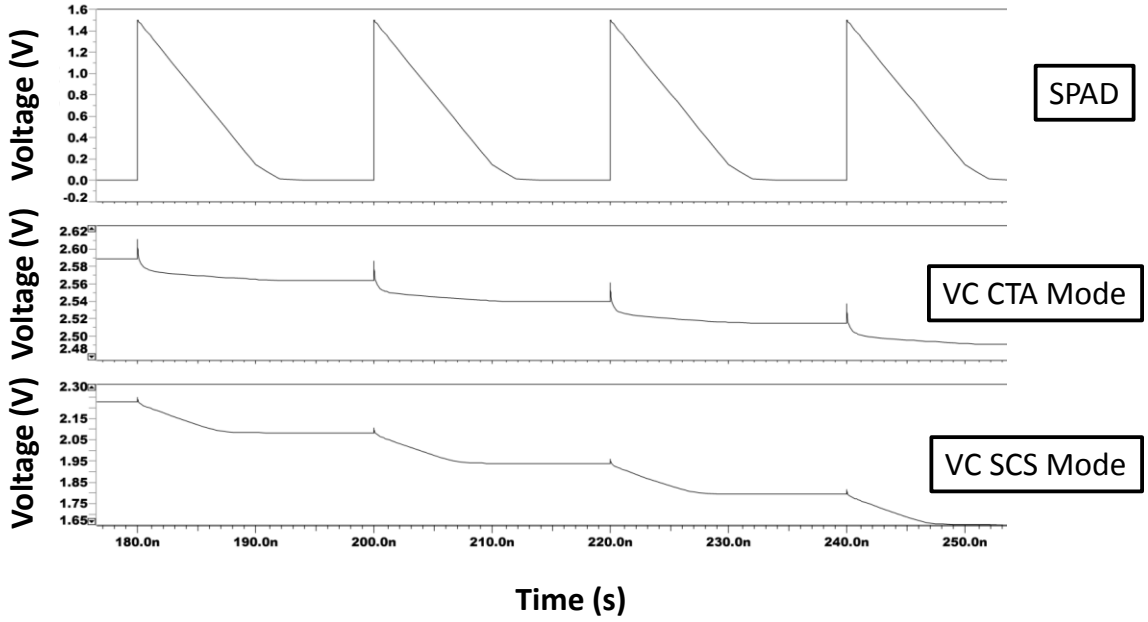


Figure 4.7. A zoom in section of the transient simulation of node V_C in both CTA and SCS modes.

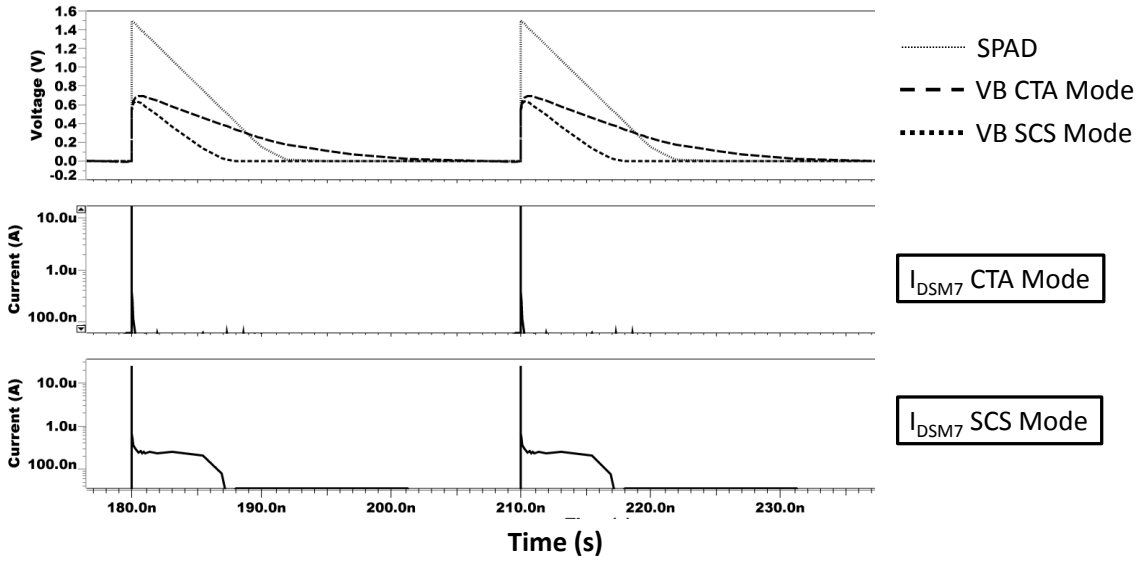


Figure 4.8. A zoom-in of the transient simulation showing both the analogue counter V_B node and the current I_{DSM7} in CTA and SCS modes.

4.2.2.2. Counter Circuit Device Sizes

Table 4.2 describes the sizes of the counter device sizes. Switches M6, M10 and source follower M9 are kept minimum sized to allow the counter M7 and M8 devices and the MOS capacitor to be sized appropriately. Increasing the length of M7 and M8 decreases the threshold variability which decreases proportionally to the square root of the device area to first order. The total capacitance C_T is calculated using the 'PLS' parasitic extraction tool to be 17fF and the parasitic capacitor C_P is calculated to be 0.1fF.

Transistor	Description	Width (µm)	Length (µm)
M6	Counter Reset Switch	0.3	0.35
M7	Counter Source Follower	0.3	1
M8	Counter Bias and Charge Drain	0.3	1
M9	APS Read Source Follower	0.3	0.35
M10	APS Read Switch	0.3	0.35
MC	Counter MOS Cap	1.25	2.74

Table 4.2. Analogue counter device sizes.

4.2.2.3. CTA Mode

The following is a description of the CTA counter modality. V_S is biased above ground, and V_G is set below the threshold voltage to keep M8 in weak inversion and above cut-off. Once the time-gate is activated, the SPAD anode is connected to the gate of M7, the counter input. The SPAD is biased above breakdown with an excess bias greater than the threshold of M7. Each SPAD avalanche event pulls V_{IN} above the V_T of M7 initiating the charge transfer operation. Charge begins flowing from the pixel capacitor to the parasitic capacitance 'C_P' at node V_B. As described in Equation 4.1, initially V_B is biased to the V_S node voltage and $V_{DS\_M8} \approx 0V$. As charge passes through M7 into node V_B, the node voltage rises due to the parasitic capacitor harging. The current through M8 (in weak inversion) is much lower than the initial current through M7 (in strong inversion) and so V_B quickly charges up to a peak value described in Equation 4.2. The two conditions of node V_B are expressed as follows:

$$V_B\,(Initial) = V_S$$

(Eq 4.1)

$$V_B\,(Peak) = V_{GS_{M7}} - V_{TM7}$$

$$= \; V_{IN} - V_B(Initial) - V_{TM7}$$

$$= \; V_{IN} - V_S - V_{TM7}$$

(Eq. 4.2)

where V_{IN} is the peak voltage of the SPAD event pulse and V_{TM7} has the conventional expression describing the body effect with V_{BULK} = 0V:

$$V_{TM7} = V_{T0} + \gamma . \left(\sqrt{\varphi_N + V_S} - \sqrt{\varphi_N}\right)$$

(Eq. 4.3)

Where γ is the body effect co-efficient and φ_N is the electro-static potential of the substrate. As the source of M7 (V_B) reaches its peak condition, V_{GS_M7} decreases and M7 moves into the sub-threshold cut-off region, halting the flow of charge and hence taking a discrete packet of charge off C_T. This dynamic operation is the fundamental difference of M7 acting as a NMOS charge transfer amplifier instead of either a continuous-time source follower or a current source switch.

After the charge transfer operation and M7 entering the cut-off region, M8 subsequently discharges the parasitic capacitor. If the discharge time of parasitic node V_B is greater than the discharge time of V_{IN} (the dead time of the SPAD), then M7 will be maintained in the sub-threshold cut-off region and the counter operates solely in CTA mode. The voltage step in CTA mode is then expressed to a first order expression as follows:

$$\Delta V_C = \left(\frac{C_P}{C_T}\right) . (V_B\ (Peak))$$

(Eq.4.4)

combining Equations 4.2 and 4.4:

$$\Delta V_C = \left(\frac{C_P}{C_T}\right) . (V_{IN} - V_S - V_{TM7})$$

(Eq.4.5)

From this expression it can be seen that the counter step size is linearly proportional to the excess bias of the SPAD and, importantly, is controllable by the external bias voltage V_S. The V_G bias voltage is used to compensate for the body effect by maintaining V_{GSM7} against V_S increases, whilst maintaining M8 in weak inversion.

In this particular implementation of the in-pixel CTA, the voltage step is proportional to the excess bias of the SPAD whereas in comparison to another recent work using a CTA based pixel counter, Panina et al. [148] rebuffer the SPAD pulse through a CMOS inverter. This re-buffering has the advantage of removing the relationship of counter step size to the SPAD excess bias. However, this is at the expense of PMOS in-pixel. If such a circuit was implemented, the step size relationship (Equation 4.5) would have the excess bias term (V_{IN}) replaced with the CMOS inverter supply voltage.

4.2.2.4. CTA Mode Variability

By the propagation of errors, the pixel to pixel variability of the counter in CTA mode can be evaluated and is shown in Eq.4.6. The expression for CTA mode variability previously published in [1] does not consider the capacitor variation whereas the expression in this work does. The V_S terminal is global to the whole pixel array, and so is assumed constant and is removed from consideration, although it will contribute temporal noise. M2 time gate variability is also not considered by assuming the SPAD excess bias (V_{EB}) is below ($V_{TimeGate} - V_{TM2}$) ≈2.4V.

$$\frac{\sigma_{\Delta V}}{\Delta V} = \sqrt{\left(\frac{\sigma_{CP}}{CP}\right)^2 + \left(\frac{\sigma_{CT}}{CT}\right)^2 + \left(\frac{\sigma_{VB}}{VB(Peak)}\right)^2}$$

(Eq. 4.6)

$$\frac{\sigma_{\Delta V}}{\Delta V} = \sqrt{\left(\frac{\sigma_{CP}}{CP}\right)^2 + \left(\frac{\sigma_{CT}}{CT}\right)^2 + \left(\frac{\sqrt{(\sigma_{VEB})^2 + (\sigma_{VTM7})^2}}{V_{EB} - V_{TM7} - V_s}\right)^2}$$

(Eq. 4.7)

As C_P, C_T and V_{TM7} variations are process dependent, from Equation 4.7 it can be noted that the voltage step non-uniformity will be non-time varying per pixel if V_{EB} remains constant. It can be deduced that the CTA pixel mode will suffer from greater PRNU with smaller voltage steps and vice versa.

4.2.2.5. SCS Mode

SCS mode is realised by biasing M8 above threshold voltage in strong inversion effectively as a current mirror. This allows M7 to function as a switched cascode device. This analysis assumes the excess bias voltage minus the M7 threshold voltage is greater than the capacitor voltage. In this condition, M7 is not cut-off and charge continues to flow from the capacitor for the duration of the SPAD dead time (τ_{SPAD}) that is above V_{TM7}. The voltage step in SCS mode is therefore expressed as a switched current:

$$\Delta V_C = \left(\frac{1}{C_T}\right) \cdot (I_{DSM8} \cdot \tau_{SPAD})$$

(Eq.4.8)

To begin expanding the SCS voltage step expression, the SPAD dead time τ_D is the sum of the avalanche and quench time (τ_Q) and the recharge time (τ_R) in the PQPR circuit:

$$\tau_D = \tau_Q + \tau_R$$

(Eq.4.9)

Expanding τ_R as the SPAD diode capacitance C_{SPAD} recharged by the current I_{DSM1} through the M1 recharge transistor with excess bias V_{EB}, the SPAD dead time is then:

$$\tau_D = \tau_Q + \frac{C_{SPAD} \cdot V_{EB}}{I_{DSM1}}$$

(Eq 4.10)

Assuming the M1 quench transistor is in saturation for the dead time period (i.e. the duration above the threshold voltage of M7) and not considering the channel length modulation effect on M1, the PQPR dead time can be expressed with VQ as a parameter:

$$\tau_D = \tau_Q + \frac{C_{SPAD} \cdot V_{EB}}{\frac{\beta_{M1}}{2} \cdot \left(V_Q - V_{TM1}\right)^2}$$

(Eq 4.11)

where β_{M1} has the conventional expression: $\beta_{M1} = K' \cdot W_{M1} / L_{M1}$

The current through M8 is of course dependent on the device drain-source voltage. Initially M8 is in saturation, and after a number of counter discharge events, the voltage on the capacitor C_T will have decreased and M8 will be in the linear region. Therefore the current through M8 is described by the two standard expressions both considering channel length modulation:

$$I_{DSM8}(Saturation) = \frac{\beta_{M8}}{2} \cdot (V_{GSM8} - V_{TM8})^2 \cdot (1 + \lambda V_{DSM8})$$

(Eq 4.12)

where $V_{DSM8} > V_{GSM8} - V_{TM8}$, $\beta_{M8} = K' \cdot W_{M8} / L_{M8}$

$$I_{DSM8}(Linear) = \beta_{M8} \cdot \left(V_{GSM8} - V_{TM8} - \frac{V_{DSM8}}{2}\right) \cdot V_{DSM8} \cdot (1 + \lambda V_{DSM8})$$

(Eq 4.13)

where $0 < V_{DSM8} < V_{GSM8} - V_{TM8}$

Considering first the saturation region, the voltage step in SCS mode is again bias controllable but with numerous difficulties in achieving this control. In contrast to CTA mode, the step size is proportional to the squared expression of excess bias minus V_S bias causing the step size to be difficult to control. Moreover, by inspection, the channel length term $(1 + \lambda V_{DSM8})$causes a non-linearity in the SCS counter response. Minimising the VS bias to lower the body effect will reduce the non-linearity, but to the detriment of step size controllability and channel leakage current.

Secondly in the linear region, as the capacitor voltage minus the source voltage V_{DSM8} ($= V_C - V_S$) falls below V_{DSM8_SAT} ($= V_{GSM8} - V_{TM8}$), the step size becomes dependent on the magnitude of the residual counter voltage (V_C). The non-linearity is evident as the step size in the SCS linear region is proportional to a non-linear expression as shown in the following equations:

$$\Delta V_C \propto \left(\frac{{V_{DSM8}}^2}{2}\right) . \ (1 + \lambda V_{DSM8}) \qquad \text{(Eq. 4.14)}$$

$$\Delta V_C \propto \frac{1}{2} \ .(V_C - V_S)^2 . \ \left(1 + \lambda (V_C - V_S)\right) \qquad \text{(Eq. 4.15)}$$

$$\Delta V_C \propto \frac{1}{2} \ .[(V_C - V_S)^2 + \lambda . (V_C - V_S)^3] \qquad \text{(Eq. 4.16)}$$

In the SCS linear region, the counter response will be non-linear in comparison to the saturation region. Moreover the MOS capacitor MC, will similarly become non-linear around $V_C \leq V_{TMC}$. Although if the counter is operated with a non-zero V_S bias, M8 will enter the linear region before the capacitor becomes non-linear. Higher V_S bias voltages reduce both the maximum counter voltage swing and the linear range of the counter.

Considering only M8 in the saturation region, by combining Equations 4.8, 4.11 and 4.12, the voltage step in SCS mode to first order is fully expressed as:

$$\Delta V_C = \left(\frac{1}{C_T}\right) . \left(\left(\frac{\beta}{2} \ . \ (V_{GSM8} - V_{TM8})^2 . (1 + \lambda V_{DSM8})\right) . \left(\tau_Q + \frac{C_{SPAD} . V_{EB}}{\frac{\beta}{2} . \left(V_Q - V_{TM1}\right)^2}\right)\right)$$

(Eq.4.17)

The parameters V_{GSM8}, V_{DSM8} and V_{TM8} are intentionally not replaced with constituent parts (e.g. $V_{GSM8} = V_{EB} - V_S$ in steady state) as each has a time varying component, i.e. during the switch on period of the current, the voltage at V_B rises to a steady state as M8 enters saturation, then falls off in line with the SPAD discharge and switches off. This time-domain switch on and off behaviour is not represented by Equation 4.17, instead it only covers the central steady-state to a first order. This SCS step equation is valid to the degree that it indicates that the controlling voltages in this mode are the excess bias V_{EB}, the quench voltage V_Q, and the bias voltages

V_S and V_G. However, none of these control voltages is linearly proportional to the counter voltage step size highlighting an inherent difficulty in controlling the SCS pixel response.

4.2.2.6. SCS Mode Variability

Applying error propagation theorem to Equation 4.8, the variation in SCS step size is defined as:

$$\frac{\sigma_{\Delta V}}{\Delta V} = \sqrt{\left(\frac{\sigma_{CT}}{CT}\right)^2 + \left(\frac{\sigma_{IDSM8}}{I_{DSM8}}\right)^2 + \left(\frac{\sigma_{\tau D}}{\tau_D}\right)^2}$$

(Eq 4.18)

Expanding the σ_{IDSM8} term, and assuming no body or channel length effects to simplify analysis, and the gate voltage is common to all pixels, the variation of current I_{DSM8} is expressed as follows:

$$\frac{\sigma_{IDSM8}}{I_{DSM8}} = \sqrt{\left(\frac{\sigma_{WM8}}{W_{M8}}\right)^2 + \left(\frac{\sigma_{LM8}}{L_{M8}}\right)^2 + \left(\frac{2\,.\,\sigma_{VTM8}}{V_{GSM8} - V_{TM8}}\right)^2}$$

(Eq 4.19)

To calculate the variation of SPAD dead time, assuming the quench time $\tau_Q \ll \tau_R$ recharge time, and the variability in excess bias is the variability of the SPAD breakdown voltage (V_{BD}), the variation in dead time has the following components:

$$\frac{\sigma_{\tau D}}{\tau_D} = \sqrt{\left(\frac{\sigma_{CSPAD}}{C_{SPAD}}\right)^2 + \left(\frac{\sigma_{VEB}}{V_{EB}}\right)^2 + \left(\frac{\sigma_{IDSM1}}{I_{DSM1}}\right)^2}$$

(Eq 4.20)

Expanding the M1 current component of the previous expression, to simplify analysis only the saturation region of the quench transistor M1 is considered as V_{DSM1_SAT} is assumed less than the threshold of M7. Also the quench voltage is removed as it is common to all pixels.

$$\frac{\sigma_{IDSM1}}{I_{DSM1}} = \sqrt{\left(\frac{\sigma_{WM1}}{W_{M1}}\right)^2 + \left(\frac{\sigma_{LM1}}{L_{M1}}\right)^2 + \left(\frac{2\,.\,\sigma_{VTM1}}{V_Q - V_{TM1}}\right)^2}$$

(Eq 4.21)

Combining the previous two equations, the SPAD dead time error expression becomes:

$$\frac{\sigma_{\tau D}}{\tau_D} = \sqrt{\left(\frac{\sigma_{CSPAD}}{C_{SPAD}}\right)^2 + \left(\frac{\sigma_{VEB}}{V_{EB}}\right)^2 + \left(\frac{\sigma_{WM8}}{W_{M8}}\right)^2 + \left(\frac{\sigma_{LM8}}{L_{M8}}\right)^2 + \left(\frac{2\,.\,\sigma_{VTM8}}{V_{GSM8} - V_{TM8}}\right)^2}$$

(Eq 4.22)

This expression proposes that SPAD dead time can be more tightly controlled by increasing the size of the recharge device to reduce variations in width, length and threshold voltage. Combining Equations 4.18, Eq.4.19, and 4.22 the SCS voltage step variability can be fully expressed by:

$$\frac{\sigma_{\Delta V}}{\Delta V} = \left[\left(\frac{\sigma_{CT}}{CT}\right)^2 + \left(\frac{\sigma_{WM8}}{W_{M8}}\right)^2 + \left(\frac{\sigma_{LM8}}{L_{M8}}\right)^2 + \left(\frac{2\,.\,\sigma_{VTM8}}{V_{GSM8} - V_{TM8}}\right)^2 + \ldots\right.$$

$$\left.\left(\frac{\sigma_{CSPAD}}{C_{SPAD}}\right)^2 + \left(\frac{\sigma_{VEB}}{V_{EB}}\right)^2 + \left(\frac{\sigma_{WM1}}{W_{M1}}\right)^2 + \left(\frac{\sigma_{LM1}}{L_{M1}}\right)^2 + \left(\frac{2\,.\,\sigma_{VTM1}}{V_Q - V_{TM1}}\right)^2\right]^{\frac{1}{2}}$$

(Eq 4.23)

By comparing Equations 4.23 and 4.7, both the SCS and CTA modes are affected by variations in capacitor C_T, excess bias and M7 threshold voltage although not equally. On top of this, the SCS mode is affected by M7 device size variation and dead time variations (namely quench transistor size and threshold voltage). For a representative example, these equations are evaluated with nominal simulation values and process parameters and shown in Figure 4.9. It is apparent that the SCS mode inherently suffers from greater variability than CTA mode limiting its usefulness to counting a small number of SPAD events.

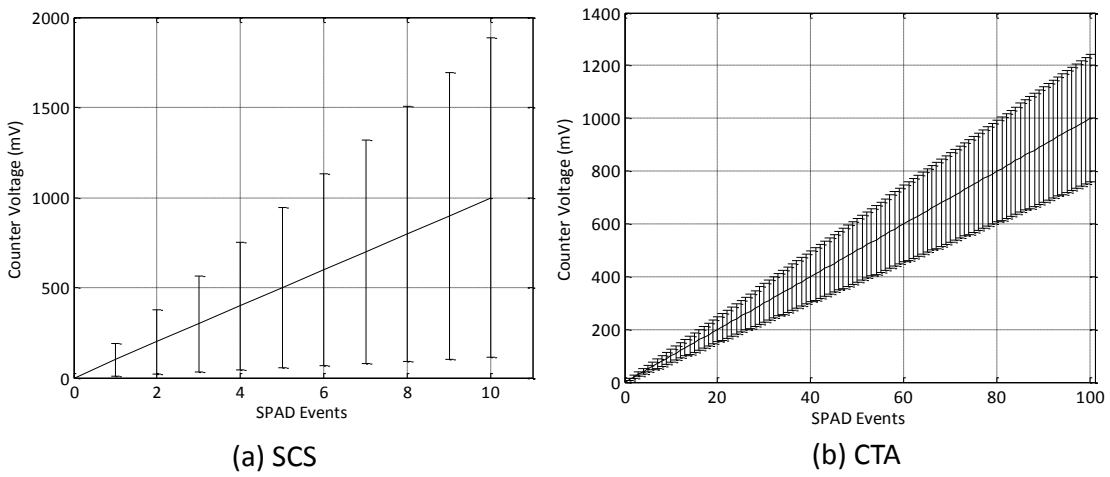


Figure 4.9. SCS versus CTA variability for nominal operating values and process parameters. Mean and ± 3 sigma error bars are shown.

4.3. Fabricated Test Array

The test array *'UNIED_TAC_PIXEL_TESTS - SPC'* was fabricated in ST Crolles in the 'IMG175' 130nm 1P4M imaging-specific process on a multi-project wafer (MPW '*I171152*'). Figure 4.10 displays a photomicrograph of the twenty two pad test structure with a silicon area of 1mm by 1mm in a ceramic 68 pin grid array (CPGA68) package. Four experimental test arrays trialling the CTA analogue counter are shown inside the IC pad ring. The right pair of arrays use the Webster DNW/P- substrate 'deep' SPAD and as mentioned at the start of this chapter have a circuit design flaw which prohibits characterisation. The left two arrays use a Richardson PW/DNW SPAD. These two arrays use the same CTA pixel circuit (and transistor sizings) but have slightly differently sized SPADs. The bottom left array attains a 12.1µm pitch (termed the 'safe' SPAD) and the top left array attains a 9.8µm pitch (termed the 'aggressive' SPAD). The top left array pixel met the sub 10µm specification by shrinking certain SPAD dimensions lower than previously trialled within the research group, hence the 'aggressive' identifier. This 'aggressive' SPAD operated with similar event rates in dark and light conditions to the 'safe' SPAD in initial tests and so the 9.8µm pitch array was selected for characterisation. The layout descriptions and pixel results described in the remainder of this chapter all come from the top left array with 9.8µm pixel pitch. The characterisation for these test arrays was performed using the characterisation and development platform PCB 1914A, detailed in Appendix 3.

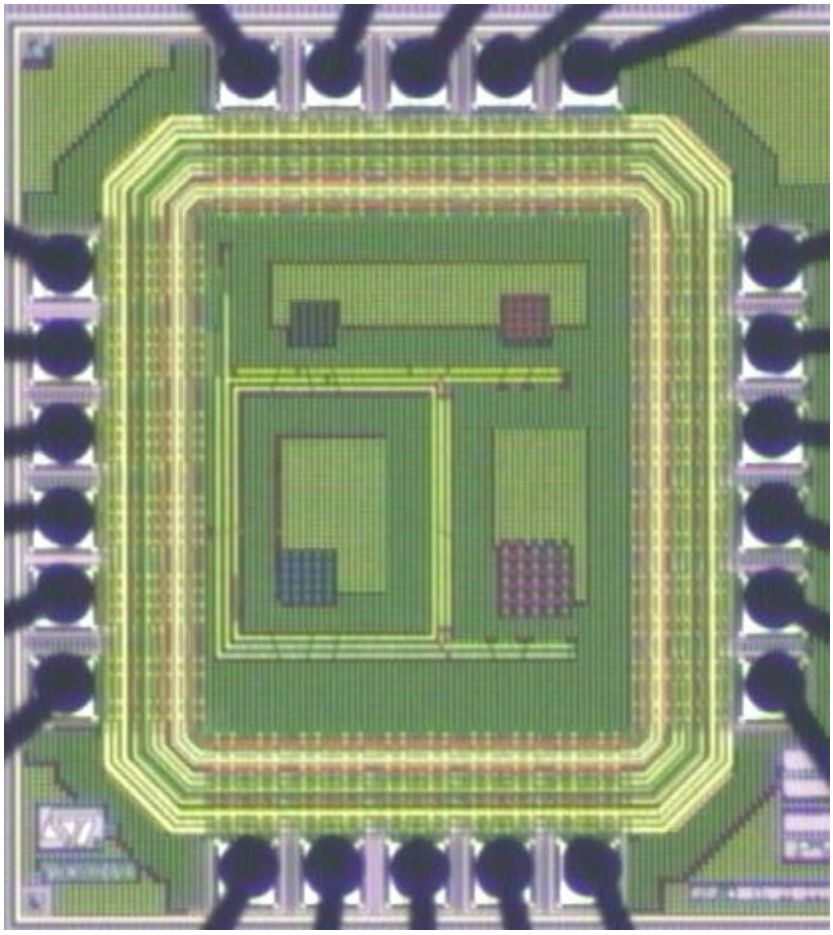


Figure 4.10. The test array IC with four 3x3 SPC test arrays. Each has a surround of dummy pixels.

4.3.1. Pixel Implementation

Figure 4.11 illustrates the schematic of the eleven NMOS-only transistor pixel. Table 4.3 provides a description of each of the pixel signals and supplies driving the SPC pixel test array.

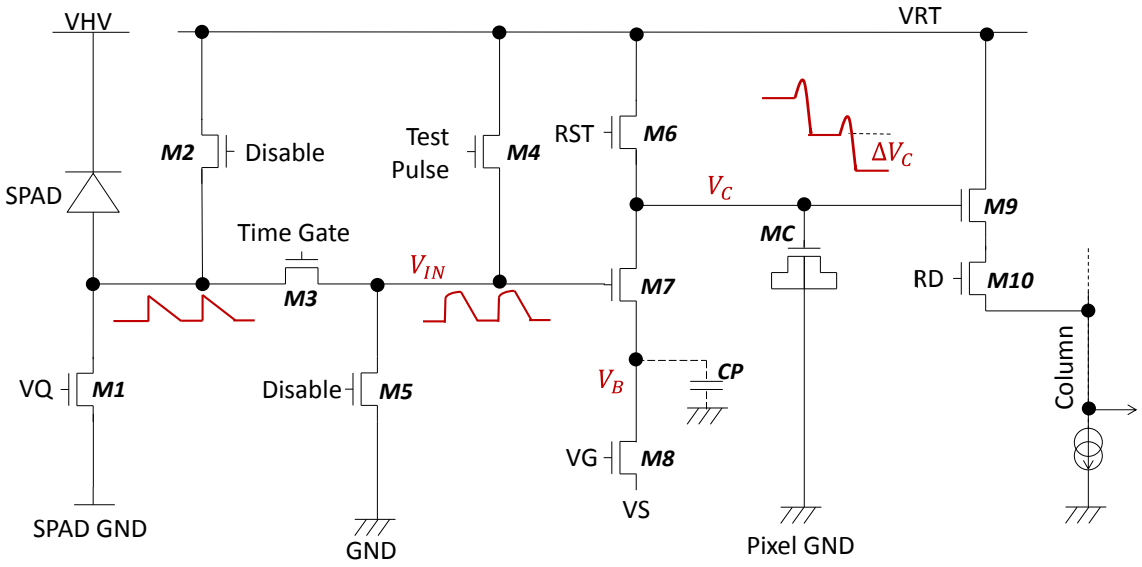


Figure 4.11. Analogue SPC pixel with 11 NMOS-only transistors and Richardson PW to DNW SPAD.

Signal	Description	Source	Typical Range (V)
GND	Substrate and quiet analogue ground, bulk connection for all pixel transistors	Chip substrate and PCB analogue ground	0
VRT	Pixel supply voltage	PCB DAC-controlled regulator	2.7
VHV	SPAD operating high voltage	PCB 0V to 30V boost circuit and driver.	14.2 to 16.2
SPAD GND	SPAD operating voltage ground		0
VQ	SPAD passive transistor quench voltage	PCB DAC	0.6 to 1.5
VG	CTA gate bias voltage	PCB DAC	0 to 3.3
VS	CTA source bias voltage	PCB DAC	0.1 to 1.0
Time Gate	Electronic shutter	Time gate row driver	3.3
Disable	Counter input pull down	Disable signal row driver	3.3
Pulse Test	Counter input pull up	PCB level shifter IC	3.6
RST	Pixel reset	PCB level shifter IC	3.6
RD	Pixel read out row select	Read signal row driver	3.3

Table 4.3. Table of supplies, input and output signals of the 11T SPC pixel and descriptions of the respective sources.

4.3.1.1. Pixel Layout

The SPAD pixel design methodology presented in Chapter 3 was developed concurrently with the design of this pixel test array. An isolated diode pixel design with surrounding 'L' shape electronics was pursued to implement this pixel circuit in a test array, as it was the state of the art at the beginning of this research (used in the 'MegaFrame' project and the doctoral research of Dr Bruce Rae, Dr David Tyndall and Dr Richard Walker [39], [40], [120], [157] and described for an analogue pixel by Dr Daniel Chitnis in [143]).

Figure 4.12 shows the 'L' shaped pixel layout achieving 9.8µm pixel pitch: 15.2 µm smaller in both X and Y directions than the state of the art SPAD-based image sensor pixel [57]. The active P-well region and the high voltage N-Well of the SPAD are highlighted and each transistor is marked. The outer white circle around the SPAD indicates the transistor P-well boundary and NMOS devices must be a further 1µm from this edge. Neighbouring pixel transistors can be seen at the edges of the layout.

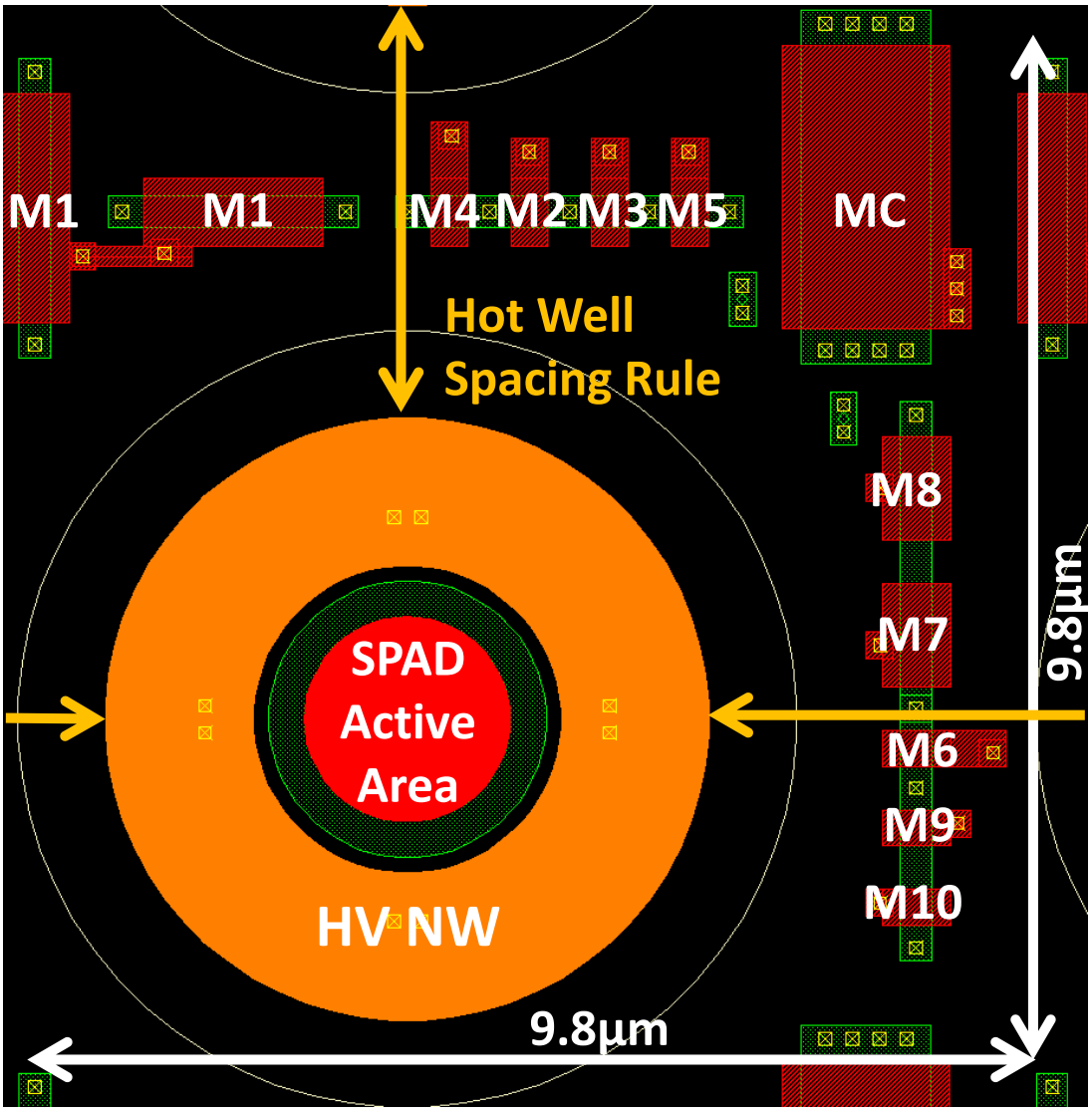


Figure 4.12. 9.8 µm SPC pixel layout: the SPAD and transistors are marked. The critical dimension for array scalability is the hot well spacing rule and is marked in orange.

The existing hot well spacing rules, marked in orange, preclude further optimisation of this pixel layout. However, customisation of the well implants and device simulation of these spacings are critical to further reduce the pixel dimensions and increase the optical sensitivity of the pixel.

The same layout is displayed in Figure 4.13, with four views incrementally adding each metal layer. Areas of metal intended to shield the analogue counter from light are highlighted. The purpose of these is to minimise photo-induced leakage in the drain and source nodes of the counter devices (MC, M6, M7, M8). Figure 4.14 shows a photomicrograph of the top left, three by three, 9.8µm pitch pixel array framed by dummy pixels.

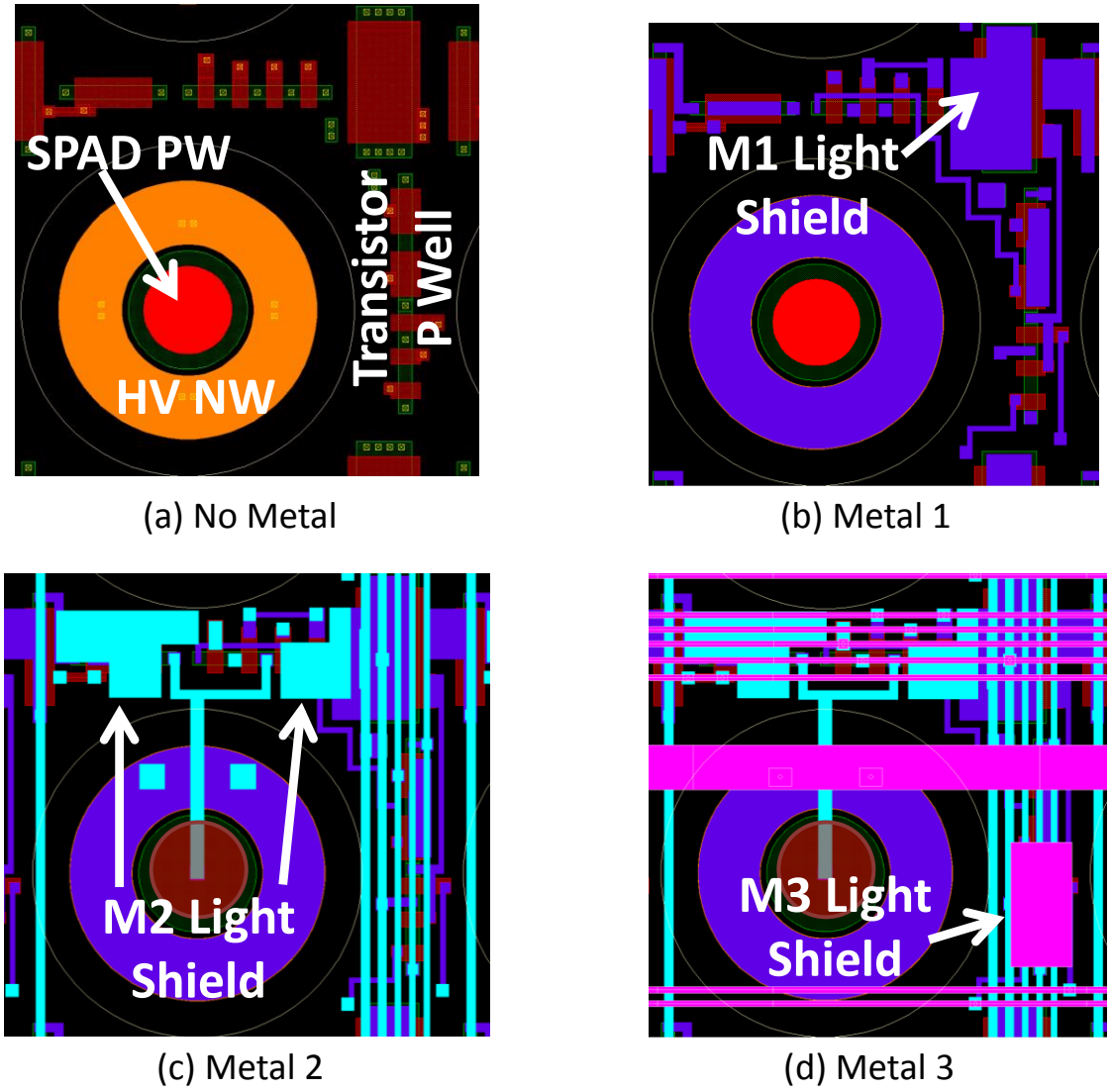


Figure 4.13. Pixel layout showing (a) no metal, and then incrementally each metal layer (b) – (d). Light shielding metal layers are indicated.

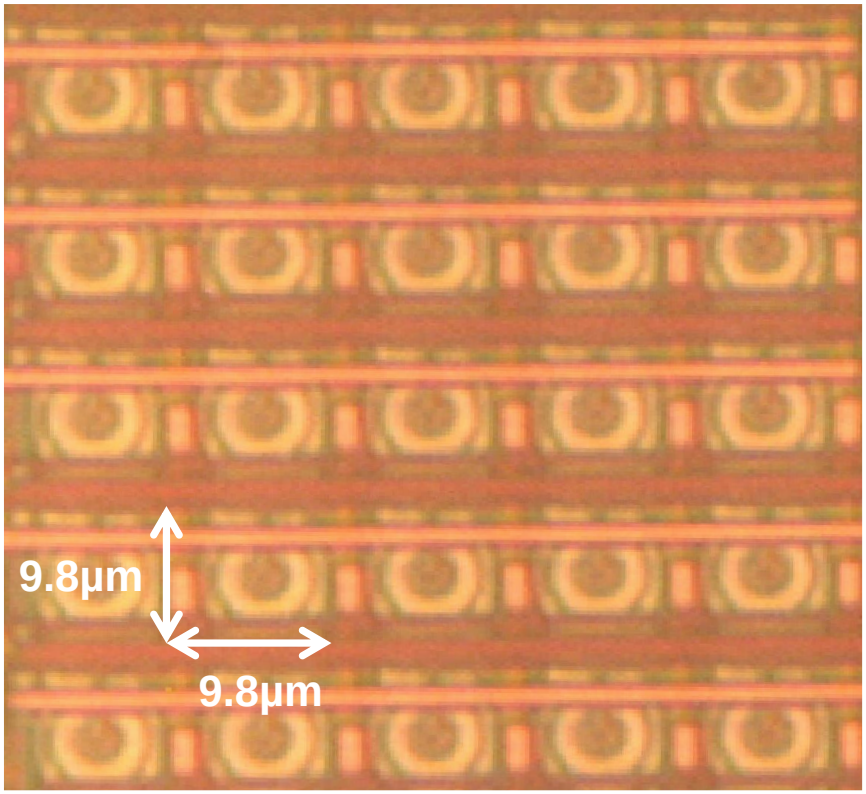


Figure 4.14. The top left 3x3 test array with a surround of dummy pixels. The 9.8µm pixel pitch is marked.

4.3.1.2. SPAD Design

The smallest isolated P-well to deep N-well SPAD reported before the beginning of this research project was published by Richardson et al. [92]. As detailed in Figure 3.12, with P-well diameter of 2µm, and outer diameter of 17.2µm it had an overall fill factor of 3.1%. Here, 9.8µm pitch is achieved whilst maintaining the fill factor, and reducing the dimensions of each constituent component. The dimensions of the designed SPAD structure are not included in this work, for confidentiality.

The quad and double row structures for the Richardson SPAD were conceived out of this initial research presented in this chapter. The quad structure was trialled in order to achieve higher fill factor and to push the existing SPAD design rules beyond the tested limits. The double row structure was based on the quad structure and was first implemented in the full imaging array described in Chapters 5 and 6 and published in [2]. Furthermore, the 'deep' SPAD equivalent of the double row structure was implemented in the 32x32 array of the 'Direct to Histogram' TDC IC published in [5] and described in Appendix 2.

4.4. Pixel Characterisation

This section details the characterisation of the hybrid-mode analogue single photon counter pixel test array. The pixel test array is characterised using the PCB 1914A (see Appendix 3 for details). The PCB was designed with two ground planes (digital and quiet analogue), 12b Linear Technology DACs (LT1446) for bias generation, 4V-maximum output low drop out (LDO) linear regulators (LT3080) for IC power supplies, and a 30V op-amp for SPAD high voltage generation. A Texas Instruments 14b differential ADC (TI ADC14L020) is operated in single ended mode with a fixed mid-rail 1.63V bias on the negative input, and the IC column output bus on the positive. An Opal Kelly 3010 FPGA board, comprising a USB interface and Xilinx Spartan 3 1000 FPGA, is used for ADC data capture and uplink to a PC. The FPGA and ADC clocks are set to 12.5MHz. MATLAB software is used for sensor and PCB control, data capture and post processing of results.

4.4.1. ADC Sampling and Noise

To characterise the temporal noise contribution of the ADC, the ADC input terminals were shorted together with a fixed mid-rail bias from a DAC. To convert ADC data number (DN) to voltage, the mid-range output code 8192 matches is set to match the measured mid rail bias voltage. Moreover, a DAC is used for calibration of the ADC matching 5500 codes to a 1.1V range. Figure 4.15 displays the measured ADC noise histogram, with a calculated standard deviation (std. dev.) of 426μV RMS. True correlated multiple sampling (CMS) is used to sample the analogue output with 4096 samples before and after integration to suppress the temporal noise of the output bus. Averaging of these samples is performed in software with floating point precision. CMS is applied in post processing in MATLAB software.

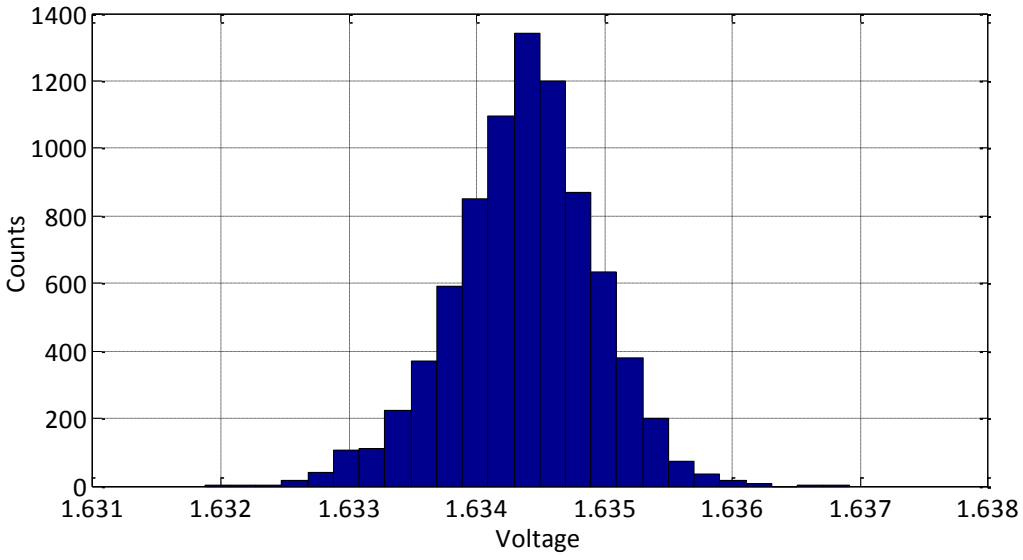


Figure 4.15. Measured ADC temporal noise histogram with 426 μV standard deviation.

4.4.2. Pulse Test Characterisation

The setup for the following experiments is as follows. The VRT supply is set to 2.7V. Both the reset and 'test pulse' transistors are NMOS switches with a switched 3.6V input. 3.6V is chosen because the maximum source voltage, that an NMOS switch in this process technology can pass from the drain, is the gate voltage minus the threshold with body effect: 3.6V – 0.9V = 2.7V hence matching the VRT supply. The FPGA timing is configured to reset the capacitor MC for a minimum of 2µs to allow complete reset settling. The time gate is disabled and the FPGA toggles the pulse test input high (and the disable input low) and vice versa to emulate SPAD pulses of 80ns duration.

The analogue counter is biased in CTA mode with M8 sub-threshold. Figure 4.16 illustrates the bias controlled sensitivity incrementing the M8 source bias 'VS' from 0 to 1V in 100mV increments. Nine pixel responses (from the 3x3 array of one part) are graphed. This experimental process is repeated to evaluate the counter sensitivity, maximum counting capability or full well, photo-response non-uniformity (PRNU), and input referred noise. Ten parts are evaluated and the statistics presented in the following eight sections represent 90 pixels in both SCS and CTA modes.

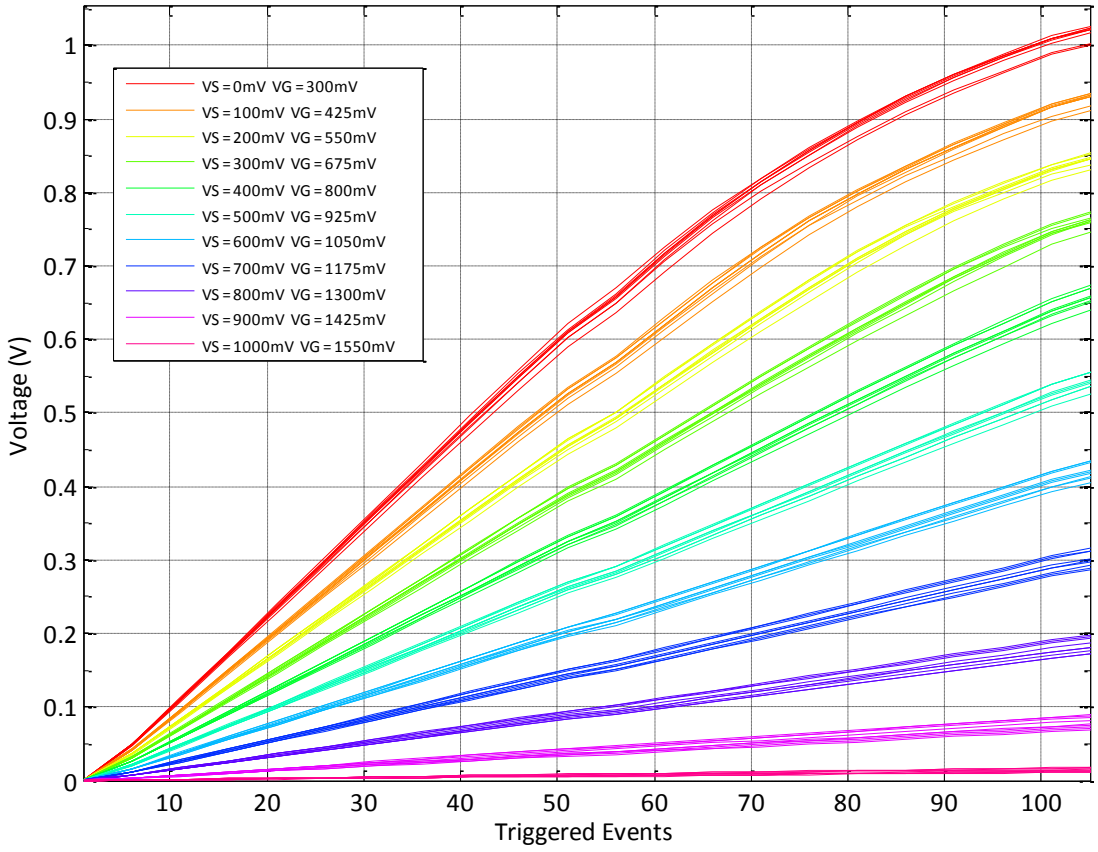


Figure 4.16. Bias controlled sensitivity of the analogue counter in CTA mode.

4.4.3. SCS Mode

The switched current source mode is evaluated using only the FPGA pulse test characterisation method. In terms of variability, this removes the contribution of SPAD dead time (by excess bias variance, quench V_{TM1} mismatch, and SPAD diode capacitance mismatch) but adds FPGA clock jitter. The source bias 'Vs' is set at 200mV for all SCS experiments to mitigate the counter leakage during capture through M7 and M8.

4.4.3.1. SCS Mode: Sensitivity

The sensitivity or counter step size was measured for varying V_G bias. The recorded voltage histogram has a set of discrete peaks for each step. The average value (centre of mass of histogram peak) is calculated per counter step. The separation or step size between each histogram peak is calculated and the mean is taken. The mean counter sensitivity in SCS mode is plotted in Figure 4.17 for incremental V_G bias. After $V_{GSM8} > 700$mV ($V_G > 900$mV) the counter fully discharges with a single trigger event.

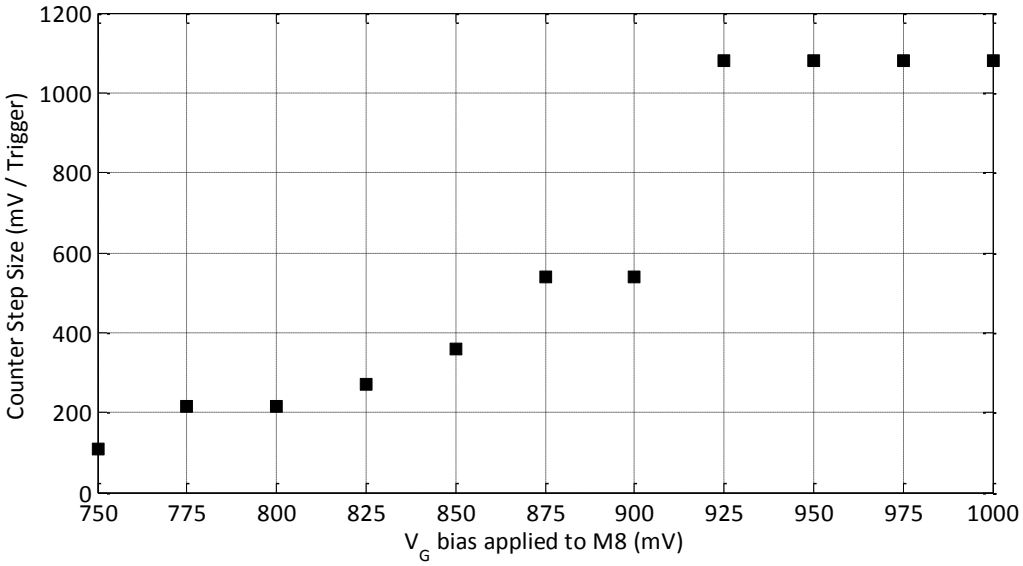


Figure 4.17. SCS mode counter sensitivity versus applied V_G bias.

4.4.3.2. SCS Mode: Full Well

The maximum counting capability or effective 'full well' of the pixel is calculated by dividing the full 1.08V range by the sensitivity. The inferred values are rounded down to the nearest integer. Figure 4.18 illustrates the 1 to 10 SPAD events maximum counting capability of the SCS mode.

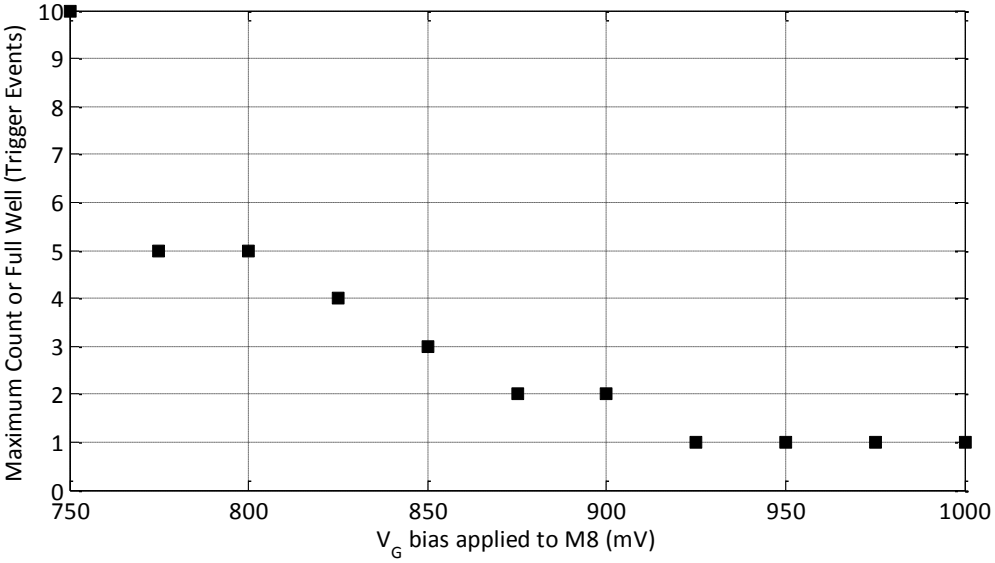


Figure 4.18. SCS mode counter equivalent full well to applied V_G bias.

4.4.3.3. SCS Mode: Non-Uniformity

The non-uniformity of the SCS pixel response, across the sampled pixels, is measured. One standard deviation of the recorded sensitivity is calculated and normalised against the counter step size recorded in the previous experiment and graphed in Figure 4.19. The photo-response non-uniformity (PRNU) is unacceptably high with the exception of the final three data points with full discharge. As high variability is evident in this SCS mode, it precludes accurate photon counting operation yet instead offers operation as a time-gated dynamic memory pixel with binary response.

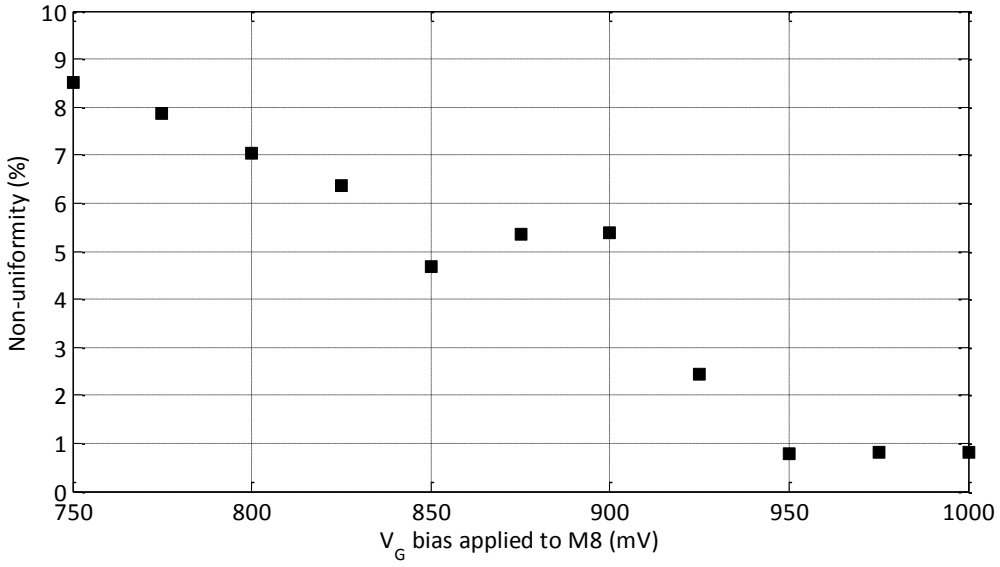


Figure 4.19. Normalised SCS mode counter non-uniformity against applied V_G bias.

4.4.3.4. SCS Mode: SPAD Input Referred Noise

For a comparison with other imaging technologies, the pixel noise is referred back by dividing by the pixel gain, to normalise the noise response in terms of photo-electrons (or SPAD counts). This is purely a mathematical exercise as fundamentally the pixel noise (mV) when input-referred to the excess bias of the SPAD (>1V) or swing of the FPGA pulse trigger (2.7V) is negligible. In Figure 4.20, the standard deviation is divided by the 2.7V trigger emulating a SPAD operating with 2.7V excess bias. The final data point value indicates 0.4 x 10^{-4} e- input referred noise in single bit dynamic memory operation.

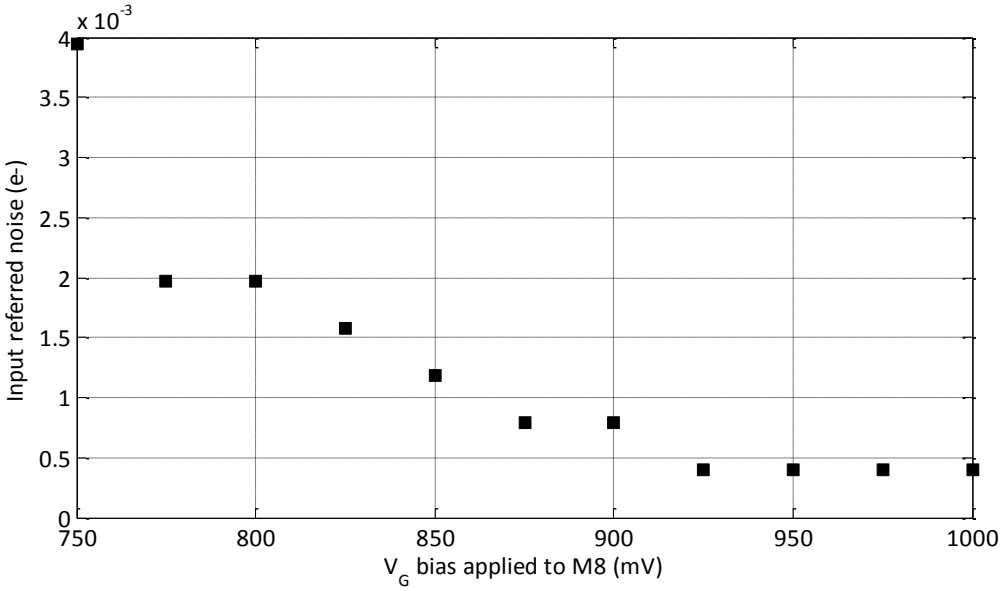


Figure 4.20. SCS mode SPAD input referred noise.

Using Fossum's conversion from input referred read noise to bit error rate (BER) from [61] is a useful metric for a binary response image sensor:

$$BER = \frac{1}{2} . erfc \left[\frac{V_n}{\sqrt{8} . s}\right]$$

(Eq 4.24)

Where s is the sensitivity (mV per SPAD event) and V_n is the voltage read noise (mV). However, any read noise below 10^{-3} e- is effectively zero bit error rate. The bit error rate conversion from 0.4 x 10^{-4} cannot be readily computed as the erfc function is asymptotically tending to zero. In Chapter 6, a more accurate method of determining the read noise and BER of a binary image sensor is presented.

4.4.4. CTA Mode

In this section the CTA mode is evaluated using the FPGA pulse test characterisation method. The following section details the CTA mode with input from the SPAD.

4.4.4.1. CTA Mode: Sensitivity

The counter sensitivity in CTA mode is captured using the same automated FPGA-based experimental method. The sensitivity is calculated as the mean separation of the peaks as per the previous SCS experiment. The CTA mode sensitivity is shown in Figure 4.21 against a linear fit of the data. The linear fit of sensitivity to applied V_S bias has the following parameters:

$$\Delta V_C = 12.54mV \ - 0.01309 \, V_S$$

(Eq 4.25)

The parasitic capacitance extraction tool reports C_P = 0.1fF and C_T = 17fF. Also, the simulator tool reports the value of V_{TM7} = 584mV in the typical process corner. Using ΔV_{IN} = 2.7V, a comparison is shown in Table 4.4, of these parameters substituted into the first order CTA equation in Eq.4.5 to the linear fit in Eq.4.25. The offset parameters show a good fit, yet, the capacitance ratios are approximately a factor of two apart. The capacitance ratio difference of two is not surprising as the tool's extraction of the diffusion parasitic is based on two separated transistors, yet in layout these devices are joined in a single strip.

	First Order Equation	Experimental Data Fit Pulse Test
Offset parameter	12.44mV	12.54mV
(C_P / C_T) or V_S gain parameter	0.00588	0.01309

Table 4.4. Comparison of linear fit and first order equation parameters.

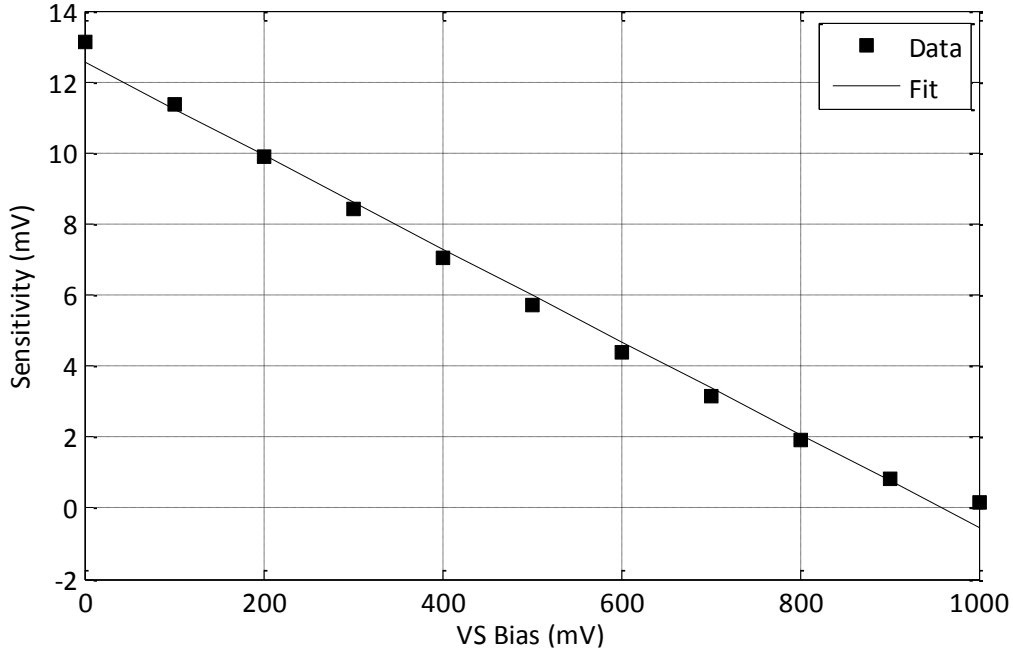


Figure 4.21. The relationship of CTA sensitivity to applied V_S bias voltage.

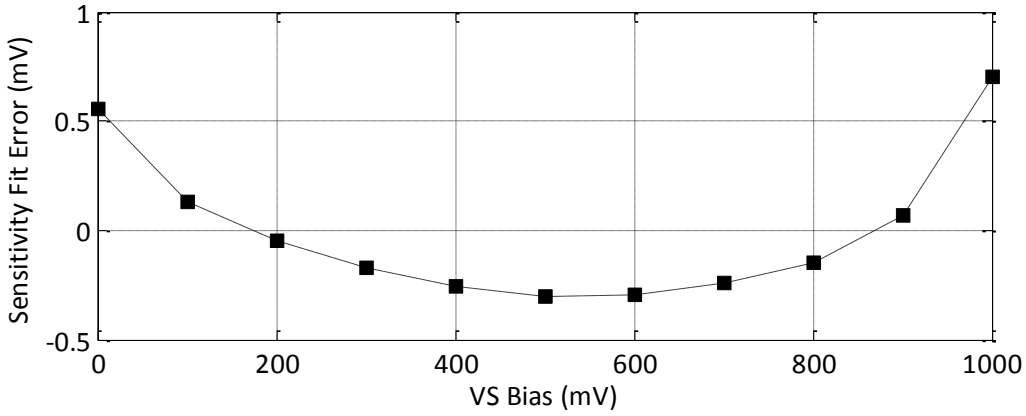


Figure 4.22. CTA sensitivity linear fit error.

Moreover, the linear fit error, demonstrated in Figure 4.22, illustrates that a second order term is not modelled in the CTA equation. There are two assumptions in the equation which offer an explanation. First, V_{TM7} is assumed constant with no body effect which is incorrect given that the CTA mode of operation relies upon M7 source degradation. Secondly, it is assumed the primary mode of operation here is solely by charge transfer. However, the counter has hybrid operation between switched current and charge transfer. If the discharge time of the M7 source node (V_B) is less than the hold or high time of the automated voltage pulse then switched current operation may have an effect.

4.4.4.2. CTA Mode: Full Well

The maximum counting capability of the pixel is measured by pulsing the test transistor M4 through a number of repetitions and recording the voltage and repetition count at which the pixels saturate. The mean of the data is calculated and plotted in Figure 4.23. This indicates that a maximum counting capability of between 80 to 400 SPAD events is achievable in the range of V_S 0 to 800mV. In the range 800mV to 1000mV, the counter can record 400 to 9000 SPAD events on average but the non-uniformity is high in this region as described in the next section.

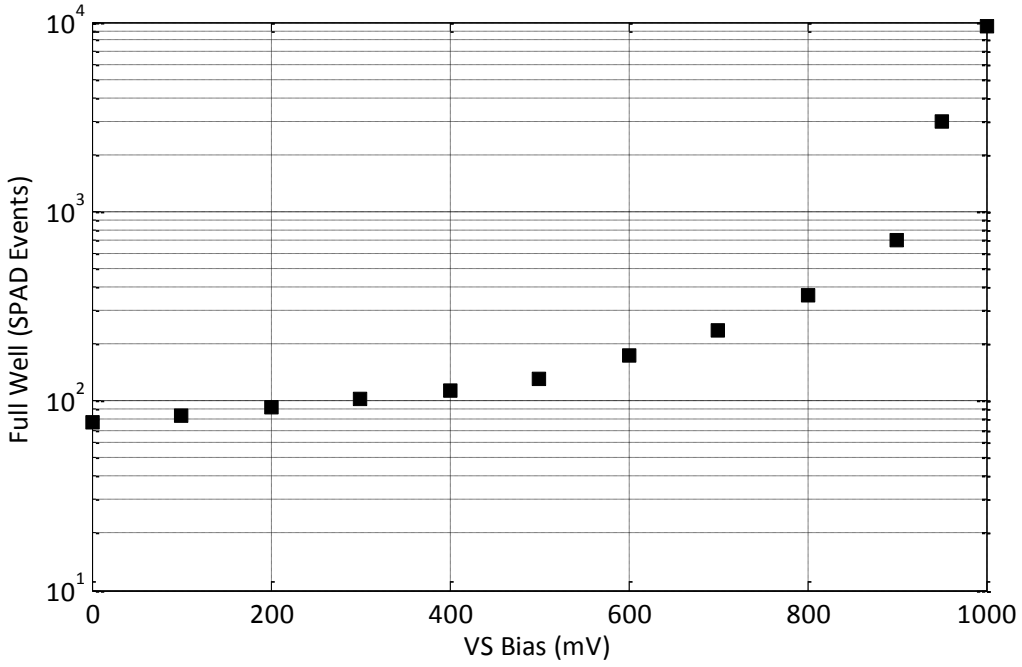


Figure 4.23. CTA mode full well or maximum counting capability versus V_S bias.

4.4.4.3. CTA Mode: Non-Uniformity

One standard deviation of the recorded full-well data is calculated and normalised against the counter step size recorded in the previous experiment. The normalised non-uniformity is graphed in Figure 4.24.

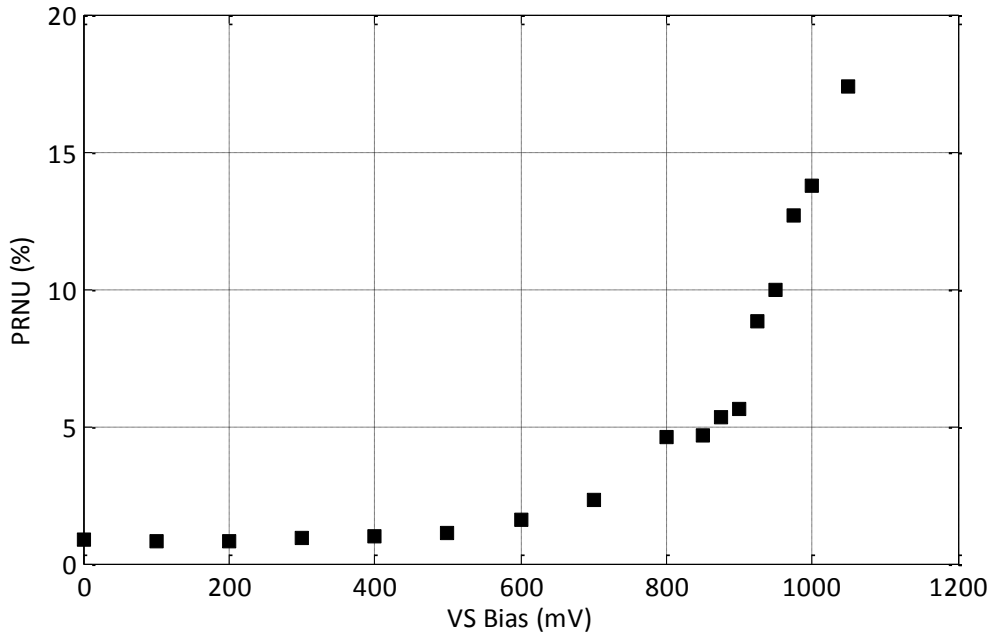


Figure 4.24. CTA mode non uniformity versus V_S bias.

4.4.4.4. CTA Mode: Input Referred Noise

The input referred noise is calculated by dividing the CTA sensitivity by the read noise. The calculated values are expressed in Figure 4.25, using one SPAD event as the equivalent of one electron for fair comparison with other imaging technologies. These data indicate that single photon counting is achievable with less than 0.3e- read noise with V_S bias less than 800mV.

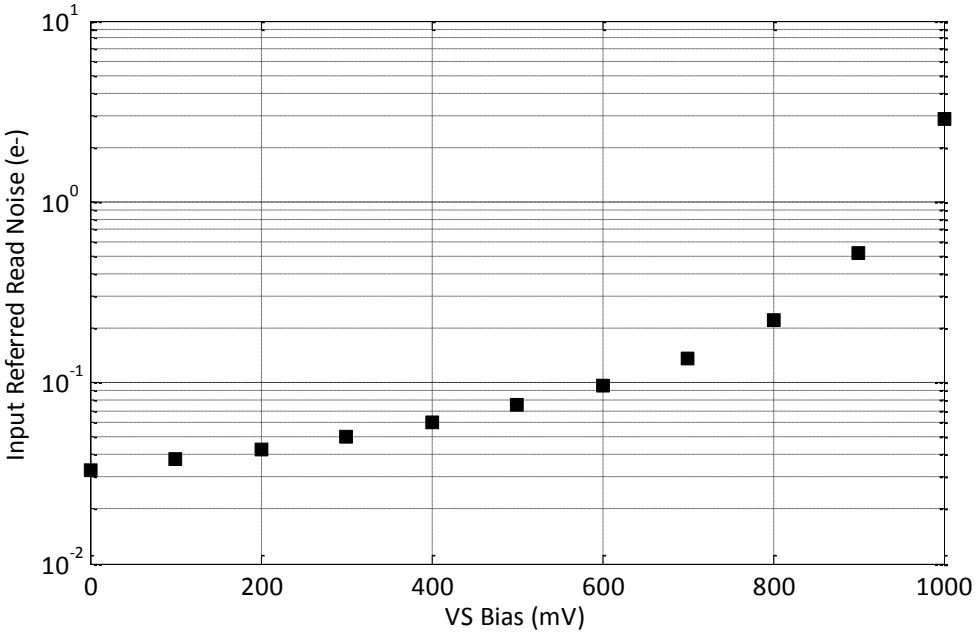


Figure 4.25. CTA mode input referred read noise.

4.4.5. Single Photon Counting

Single photon counting is achieved with the counter biased for CTA operation. The SPAD is enabled and biased at 2.7V excess bias above breakdown. Figure 4.26 illustrates an example of the output of one pixel recorded with 1,000 exposures or repetitions with 30µs integration time. The discrete peaks under a classical Poisson distribution are clearly evident indicating the photon counting in this example is truly shot noise limited. For an image sensor pixel this is not a typical response, it is more in common with the processed output of analogue SiPM or PMT, see [43] for a range of photon counting examples. Figure 4.27 illustrates a set of single photon counting histograms with 25,000 exposures showing bias controlled sensitivity with the single photon peaks becoming closer together for increased VS bias. The peak separation is calculated and these sensitivity data are graphed with a fit of the experimental values alongside the automated pulse test data from Figure 4.21 showing good alignment between the two data sets in Figure 4.27. Table 4.5 presents the extracted fit parameters of the single photon counting sensitivity, again indicating that the previous electrical pulse test experiments are in line with these single photon counting experiments.

	First Order Equation	Experimental Data Fit Pulse Test	Experimental Data Fit SPC
Offset parameter	12.44mV	12.54mV	12.62mV
(C_P / C_T) or V_S gain parameter	0.00588	0.01309	0.01383

Table 4.5. Comparison of linear fit parameters from pulse test and SPC data, and first order equation parameters.

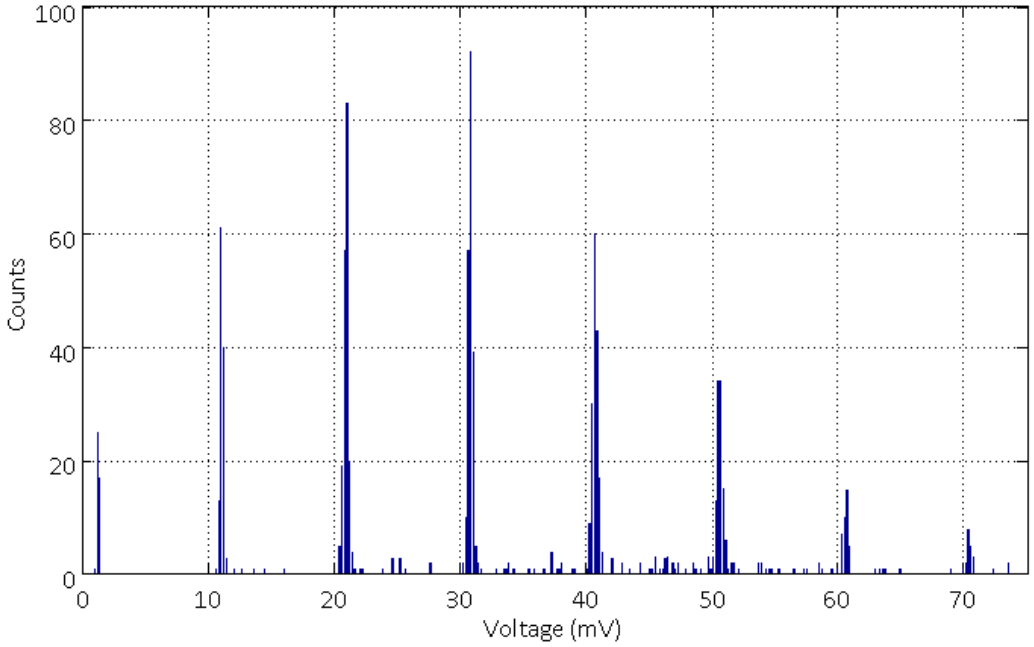


Figure 4.26. Single photon counting histogram from one pixel output.

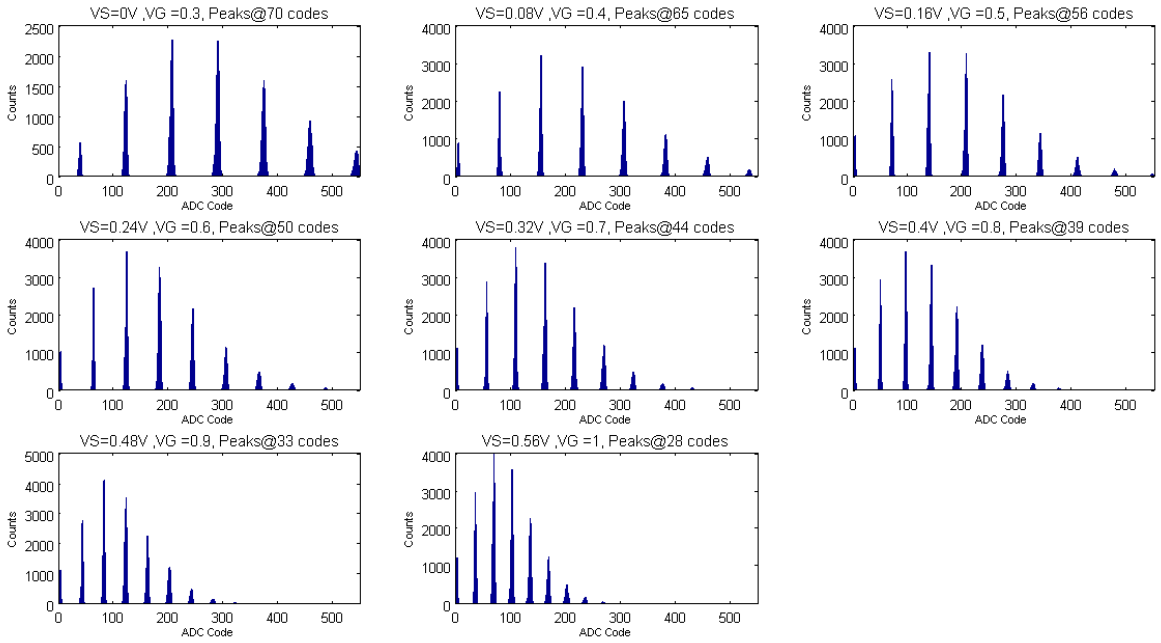


Figure 4.27. Single photon counting histograms with increasing V_S bias from top left to bottom right.

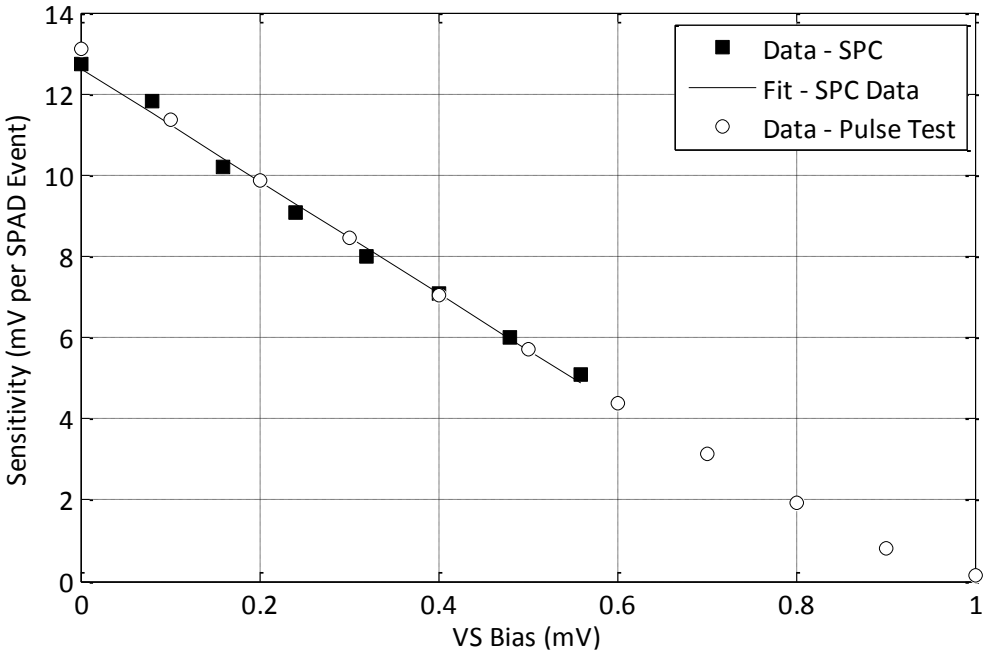


Figure 4.28. Measured single photon counting sensitivity combined with the measured electrical pulse test data.

4.4.6. Counter Noise and Distortion Measurement

A set of single photon pixel exposures are performed to evaluate the counter performance with increasing exposure time. These are illustrated in Figure 4.29(a) from left to right in 8µs increments from 8µs to 200µs. Discrete peaks are evident in the first row of captures (8µs to 40µs) but the peaks broaden and merge by the final row (168µs to 200µs). Figure 4.29(b) combines these 25 experiments into a single histogram. The 'in-filling' of counter voltages between the discrete spikes is attributed to noise and a distortion mechanism in the pixel. If the SPAD pulses before the CTA discharge node (V_B) has fully discharged then the counter step is proportional to the residual voltage. This causes subsequent voltage steps to continue from this less than full or fractional step.

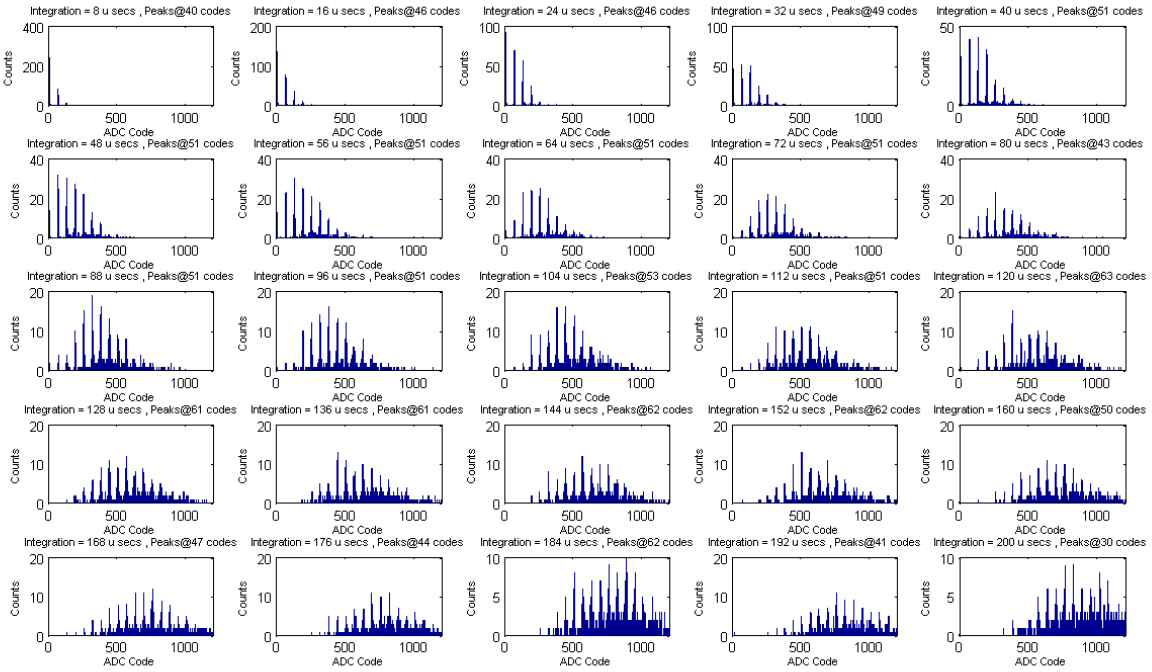


Figure 4.29. Single photon counting histograms (a) Increasing exposure time in 8µs increments. (b) Combined data in one histogram.

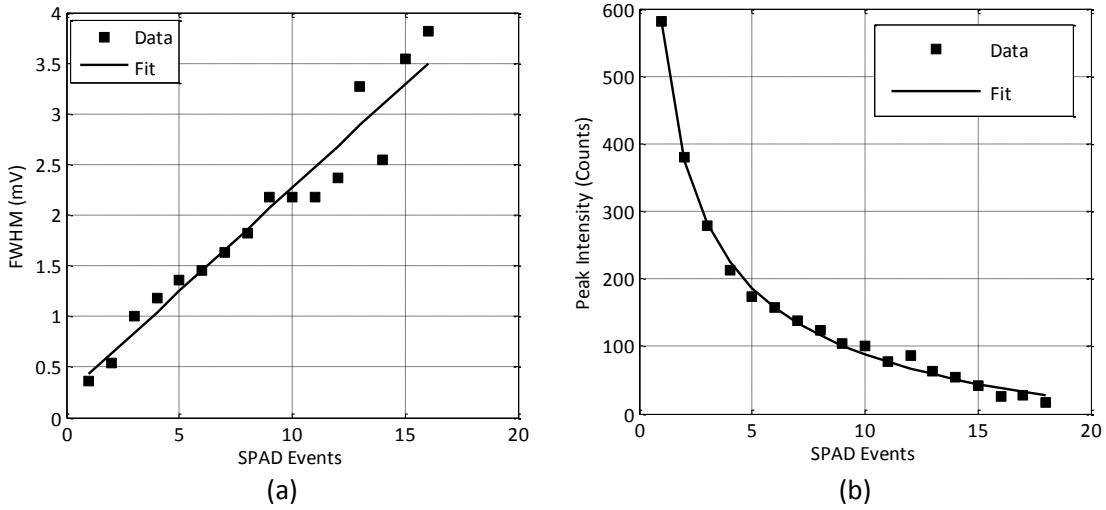


Figure 4.30. (a) Peak FWHM versus peak index. (b) Extracted peak intensity versus peak index.

	FWHM	RMS
Offset parameter	204.73μV	95.51μV
Gain parameter	224.91μV	86.94μV

Table 4.6. Peak FWHM data fit parameters.

It is evident that the single photon counting peaks decrease in height (or counts) and increase in width (or variance) for greater numbers of cumulative SPAD events. Figure 4.30(a) plots the increasing peak FWHM against increasing number of SPAD events extracted from the combined histogram data shown in Figure 4.29(b). Figure 4.30(b) shows the decrease in peak height for incremental SPAD events, if there was no distortion mechanism there would be no decrease in peak height and this graph would be a straight line. The linear fit of the FWHM data provides two parameters, described in Table 4.6, which can provide information on this distortion mechanism.

The non-zero gain parameter indicates that there is a cumulative noise and distortion mechanism of 87μV RMS noise accumulation or pulse broadening for every SPAD event. Two possible sources of this distortion are proposed; first there is an accumulation of noise on each counter SPAD event. Secondly, as previously discussed the 'in-filling' distortion causes fractional steps. As these data are for a single pixel only, it cannot be attributed to V_{TM7} mismatch, which would be seen if multiple pixels were combined. Moreover, the non-zero offset parameter may indicate that the true CMS timing and data capture does not fully cancel the reset noise.

The cumulative noise contribution to each CTA voltage step is attributed to four sources: the Johnson noise through both channels of M7 and M8 acting on C_P, and the flicker and temporal noise (voltage fluctuation) of the VS bias, and the variation in SPAD peak excess bias. The VB node will assumedly have a high degree of

kT/C noise due to the small parasitic capacitance of C_P (approximately 2mV RMS). However, the kT/C noise contribution of the VB node on the VC node will be divided by the CTA capacitor ratio. Flicker noise will contribute proportionally to the exposure or integration time. To minimise temporal noise and achieve stability on the VS bias supply requires adequate decoupling and low impendence routing to each pixel. The peak excess bias of the SPAD is assumed constant from avalanche to avalanche, however, there will be temporal noise on the SPAD cathode high voltage supply which will manifest itself as cumulative noise per SPAD event. Although it will be present for a few pixels within a large distribution, RTS noise from M7 or M8 is not considered.

4.4.7. Counter Distortion Simulation and Noise Modelling

A simulation of the CTA demonstrates the existence of a counter distortion due to the imperfect reset of the parasitic node for short inter-arrival times of SPAD events. The counter is simulated in CTA mode ($V_S = 0.1V$, $V_G=0.5V$) with a triangular wave representing a SPAD input of 1.5V excess bias with 5ns dead time. The inter-arrival time between two SPAD pulses is incrementally swept. As illustrated in Figure 4.31, on assertion of the second SPAD pulse there is a residual voltage remaining on node V_B if the parasitic capacitance has not fully discharged. The counter voltage step, occurring from this second SPAD pulse, is then distorted to a lower value dependent on the residual V_B voltage.

Figure 4.32 shows calculated analogue 'spice' simulation results of the relationship of counter voltage step and V_B residual voltage to SPAD pulse inter-arrival time. The 'knee' in the voltage step curve at 100ns represents the total discharge time of the V_B parasitic node. This clearly indicates that 'imperfect CTA reset' distortion can occur when the SPAD dead time is less than the parasitic discharge time. As the distortion has a photonic inter-arrival time dependence then it follows that it has a matching relationship with light level. For low light levels with photon arrivals much greater than the V_B node discharge time the effect will be negligible. Yet for higher light levels with higher SPAD event rates with inter-pulse time less than the V_B node discharge the effect becomes apparent. For an image, this distortion mechanism may be acceptable as it may occur only a small number of times and the final offset may not be significant. Nonetheless, for accurate photon counting this is not desirable, but of course it is an unavoidable effect particular to the chosen CTA design.

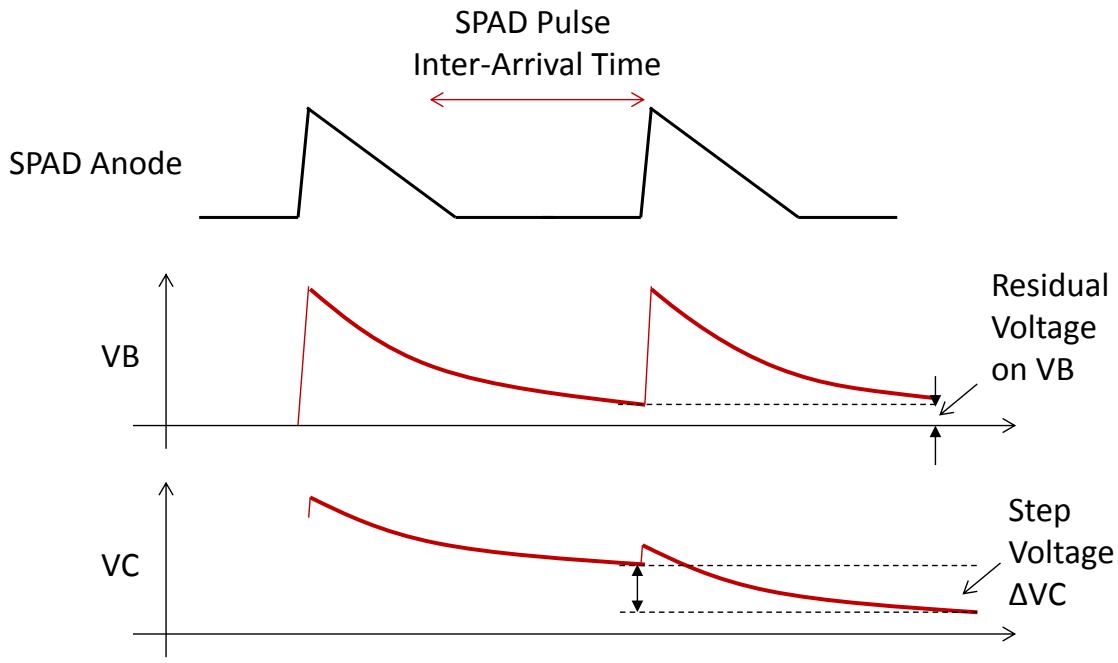


Figure 4.31. Analogue counter in-fill distortion illustration.

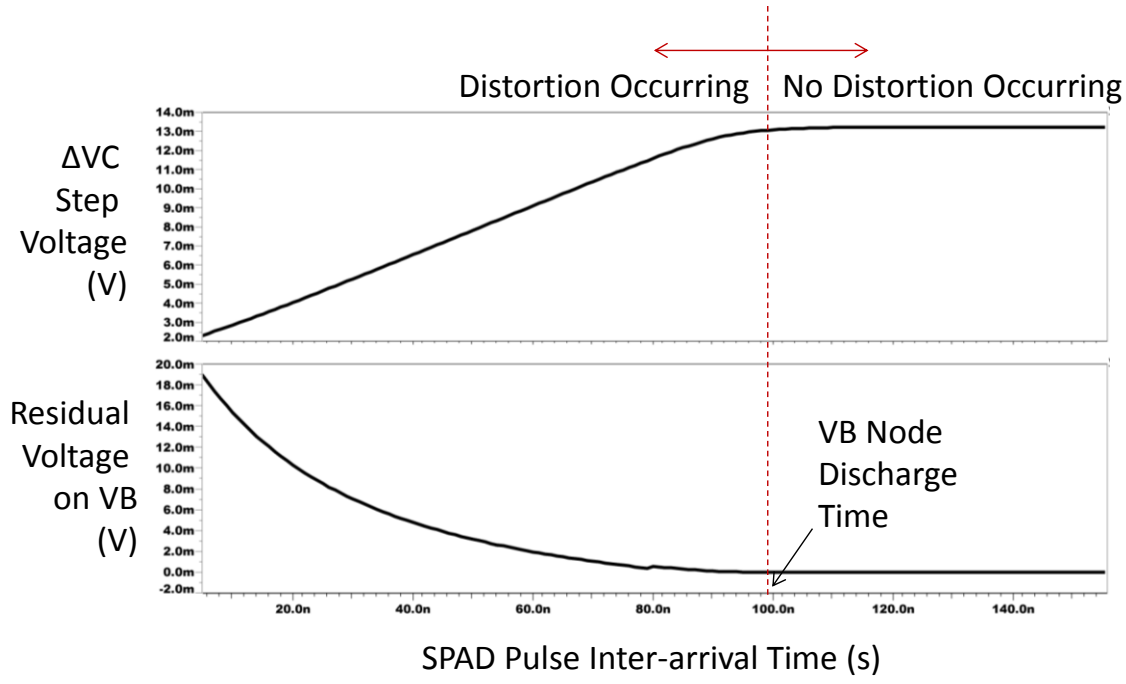


Figure 4.32. Analogue counter in-fill distortion relationship to photon inter-arrival time.

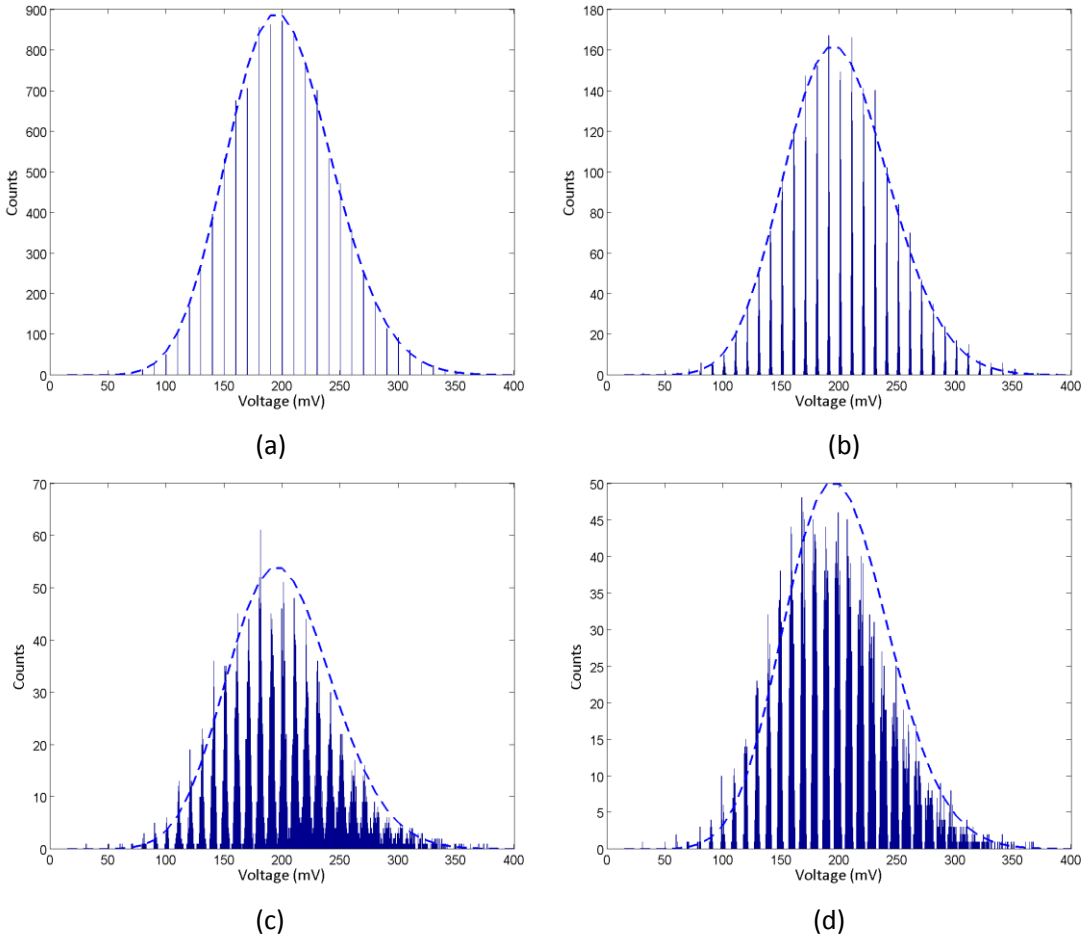


Figure 4.33. Modelling of analogue counter noise and distortion effects. Ideal Poissonian curves are overlaid with a dotted line. (a) Ideal Poissonian (b) Ideal Poissonian with added read noise. (c) Read noise and cumulative counter step noise. (d) Read noise and imperfect CTA reset distortion.

A statistical model of cumulative noise and in-filling distortion reveals that the results are a good match to the single photon counting histograms obtained. Figure 4.33(a) demonstrates a histogram of a Poissonian distribution of 10,000 photons captured using an ideal noiseless single photon counting circuit with 10mV steps. Figure 4.33(b) shows the pulse broadening effect of adding read noise (matching the 426μV RMS read noise of these experiments) where the spikes indicating single photons remain clearly visible with no deviation from the ideal Poissonian dotted line observed. Figure 4.33(c) goes further by scaling each photon count by a normal distribution representing the cumulative noise per step (86.9μV RMS) and it can be seen that there is a distortion away from the ideal Poissonian distribution. The broadening and 'in-filling' (or merging) of the later peaks (>200mV) is visible. Figure 4.33(d) simulates read noise and the contribution of 2.5% imperfect CTA reset distortion with uniform distribution of the step size for the 2.5% distorted analogue counts. The separation between the peaks diminishes.

Figure 4.34 illustrates the modelling simulation result combining the contributions of read noise, cumulative noise per step and imperfect CTA reset distortion which may be compared to the SPC graphs in Figure 4.29(a). The deviation from the ideal dotted line Poissonian curve is apparent. The conclusion of Chapter 6 provides steps to reduce the cumulative noise in the CTA-based analogue counter.

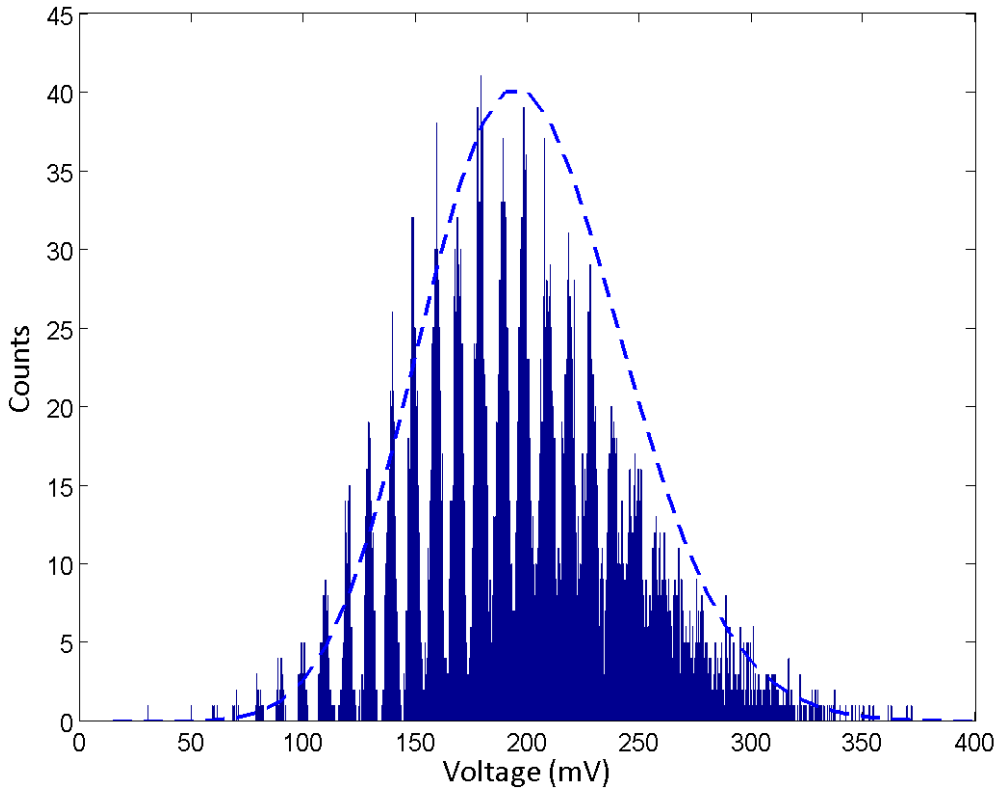


Figure 4.34. Simulation of SPAD-based CTA counter with 426µV RMS read noise, 86.9µV RMS cumulative noise per SPAD event and 2.5% imperfect CTA reset distortion.

4.5. Summary and Discussion

This chapter details an eleven NMOS transistor time-gated analogue counter circuit for single photon counting. The test array pixel is a proof of the concept that a SPAD-based pixel using an analogue approach can be minimised in pitch below 10μm without losing fill-factor of previous all-digital SPAD pixels.

The time-gating performance is not explored using this test structure as no delay control or time window generation is included on-chip. Qualitatively the short time gate was judged to be operational as single photon counting was captured using a single clock period (100ns) exposure window. Chapters 5 & 6 present the design and performance of time-gating for the image sensor.

The proposed counter architecture has hybrid operation between switched current and CTA, controlled by the V_G gate bias of the tail device of the counter circuit. The source terminal bias V_S is utilised to linearly control the counter step size (or pixel sensitivity) in CTA mode. Whereas in SCS mode, the sensitivity has a non-linear relationship to the source bias and other biases and therefore is difficult to control. Table 4.7 summarises the two conditions of the two modes of the hybrid analogue counter.

	Switched Current Source	**Charge Transfer Amplifier**
M7 Operation Mode	Source Follower	Charge Transfer
Discharge Time of Cp Node	$\tau_{VB} = \tau_{SPAD}$	$\tau_{VB} \neq \tau_{SPAD}$
M8 Region of Operation	Above Threshold, Moderate to Strong Inversion	Sub Threshold, Weak Inversion
V_G Bias	$\geq V_{TM8}$	$< V_{TM8}$
V_S Controlled Sensitivity	Non-Linear	Linear

Table 4.7. Defining characteristics of the two modes of the hybrid analogue counter circuit.

The results from this test pixel are tabulated against state of the art analogue single photon counting pixels, in Table 4.8. The CTA results are in line with the other CTA works in terms of counter step, non-uniformity and read noise. On the other hand, the CTA full well is higher by 100 counts than other works and can be increased at the cost of higher non-uniformity. The SCS operation has comparable PRNU to [144] but lower effective full well. SCS in comparison to CTA is useful to provide the large voltage step required for binary operation triggered by a single SPAD event.

In evaluation of the counter noise and distortion, Figure 4.32(b) exhibits what would be achievable with a charge-accumulation SPC pixel with very low (<0.15e-) read noise. The integrated noise visible in Figure 4.32(c) indicates the achievable SPC performance with any analogue counter structure cumulatively adding noise per step.

Author	This Work				Stoppa	Chitnis	Pancheri		Panina
Reference					[141]	[143]	[144]	[146]	[148]
Year Published	2013				2009	2010	2013	2013	2013
Counter Architecture	CTA V_S= 0V	CTA V_S= 0.8V	SCS V_G= 0.75V	SCS V_G> 1V	SCS	SCS	SCS	CTA	CTA
Counter Step per SPAD Event	13.1 mV	2.1 mV	108 mV	>1V	38mV *	140mV to 2mV†	10mV	15.6mV to 47.8mV	4 to 8mV
Effective Full Well (SPAD Events)	80	360	10	1	44	14 - 980†	140*	89 to 29	128 to 256
Read Noise	0.03e-	0.22e-	0.004 e-	3 x 10^{-4}e-	1.3LSB	NR	0.13e-	0.05 to 0.016e-	NR
PRNU	0.9%	4.6%	8.5%	<1%	3 LSB	NR	11%	1.9%	4%

Table 4.8. A comparison table of SPAD-based single photon counting analogue pixels (Key: NR = Not Reported, * Estimated, † Logarithmic counter).

The pitch is commensurate with low-light single photon imaging technologies such as EMCCD and photo-cathode intensified CCDs (8 to 16µm) and sCMOS (6.5µm) and with state of the art TOF pixels (10µm). The full well of 100's of counts in CTA mode is equivalent to the full well of EMCCD, although with higher PRNU. Yet the fill factor remains considerably lower than these other devices (>60%), and this would hamper photon-starved image capture and TOF imaging performance.

Pitch reduction and fill factor improvement are two clear targets to ensure the highest performance for an image sensor based on this pixel test structure. The SPAD guard ring dominates 87% of the pixel area with transistors occupying approximately 10% and the SPAD active area 3%. The layout of the counter and front end circuit must be optimised further to decrease the transistor area. The NMOS capacitor MC is the largest contributor; combined use of metal capacitors and poly capacitors would allow for layout optimisation and area reduction. Alternative pixel layout strategies, based on well-sharing and proposed in Chapter 3, are exploited to address the SPAD guard ring and the low fill factor in a revision of this pixel presented in the next chapter.

5. A CMOS SPAD-based SPC Image Sensor

5.1. Introduction

5.1.1. Fabricated IC

Building on the conclusions of Chapter 3 and the pixel test array described in Chapter 4, a SPAD-based image sensor was designed and manufactured to be a proof of the concept that SPADs and compact electronics could be exploited to create a high resolution time-gated photon counting image sensor. A photomicrograph of the sensor is shown in Figure 5.1. The image sensor, measuring 3.4 x 3.1mm, is fabricated in STMicroelectronics' 130nm front end/90nm back end imaging low voltage CMOS (1 poly and 4 metal) process. There are no extra implants or custom masks required for the CMOS SPAD devices. The main imaging array is in a 4:3 horizontal to vertical ratio consisting of 320 by 240 pixels at 'QVGA' graphics display resolution. The array measures 2.56mm by 1.92mm with a 3.2mm diagonal; equivalent of a 1/5.3" optical format.

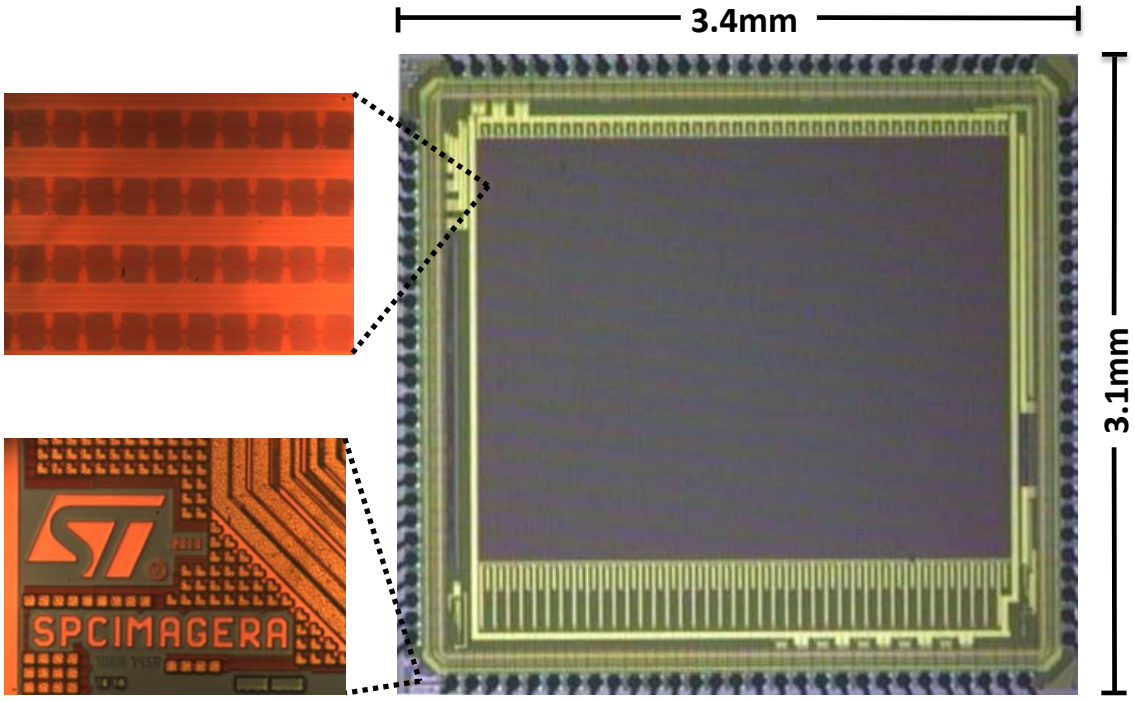


Figure 5.1. CMOS SPAD-based single photon counting image sensor fabricated in STMicroelectronics 130nm imaging process with highlights of the logo and line name (below) and the pixel array (above).

Two revisions of the sensor have been manufactured. The original revision 'AA' has time gate drivers from both sides of the array with a split in the centre, whereas the second metal fix revision 'AB' has time gate drivers from only one side with no array split. There is also a slight modification to the digital readout serialiser between the two chip revisions. Both the 'AA' design and the 'AB' modifications are documented. This chapter provides an overview of the sensor, the sensor timing and the control firmware. The sensor is detailed in terms of the pixel design and layout, the readout design and the overall sensor architecture. The data processing, sensor timing and system control are also described in this chapter, which are handled by Verilog-based firmware on a Xilinx FPGA.

5.1.2. Specifications

The aim of this research is to demonstrate the proof of principle of a high resolution time-gated SPAD-based image sensor design. The primary goal was to achieve a compact pixel with in-pixel gated analogue counting operation with reduced pixel pitch of prior SPAD-based image sensors whilst improving fill factor. A further goal was to explore single bit memory operation with fast digital readout permitting digital image oversampling. The initial project specifications, for the sensor envisaged to achieve these goals, are presented in Table 5.1.

Specification Parameter	Value
Array Size	320 x 240
Pixel Pitch	Minimum <20µm Ideally <10µm
Fill Factor	Minimum >1% Ideally >20%
Silicon Area	< 5mm x 5mm
Frame Rate	>30fps
TOF Distance Range	10cm to 150cm
TOF Error	<1cm at 150cm

Table 5.1. Initial specification table of the SPAD-based time-gated SPC imager.

The sub-20µm minimum target pixel pitch was an incremental decrease from the SiPM array presented in the doctoral work of research group member Dr David Tyndall. A maximum of 25mm^2 silicon area is stipulated in the Research and Collaboration Agreement (RCA) document between ST and the University for this sensor. A 320x240 array using this maximum area would permit a pixel pitch of 15.6µm. The 5µm pitch of Prof. Robert Henderson's SiPM test structure (ST's IMG140 90nm front end/65nm back end process) and the 11.6µm pitch of Dr Eric Webster's SiPM test structure (ST's IMG175 130nm process) indicated that improvement on that figure could be made [48], [105]. Thus, a sub-10µm pitch specification was proposed as an ideal target. An intended demonstration for the sensor is an ITOF 3D vision sensor operating in a mobile phone, tablet, laptop or desktop computer. These applications are envisaged to have a maximum distance of

1.5m from the device user or target to the ToF camera. This would imply an equivalent maximum temporal dynamic range of 10ns. Two further specifications were to provide 'clean handling' of out of range targets in a scene, and to implement an ambient rejection scheme. The frame rate specification was chosen to match commercial CISs.

5.2. Image Sensor Description

To realise these specifications, a quarter video graphics array (QVGA) of 320 columns by 240 rows of time gated single photon counting pixels was designed. Further to the main imaging array an extra 16 rows consisting of three sets of test pixels were placed above and below the main array. All 256 rows of pixels conform to the 8µm pitch and use the same SPAD. The pixel array comprises:

- 240 rows of the main single photon counting imaging array.
- Single dark row below the main array, shielded by top metal.
- 8 rows above the main array, of sample and hold TAC pixels test structures.
- 7 rows above the main array of a charge pump counter test structure.

The design and layout of the latter two test arrays was undertaken by Luca Parmesan as part of his main doctoral research [7] and as such is not further described in this thesis.

The block diagram of the image sensor is shown in Figure 5.2. The analogue single photon counting pixel outputs are read out via a conventional CIS active pixel sensor (APS) architecture. Rows are sampled sequentially into column parallel CDS sample and hold stages. Each of these CDS column buffers is scanned out using a single channel analogue bus through two single-ended op-amp buffers to an off-chip differential ADC. Further to this, as shown at the top of Figure 5.2, a fast single bit digital readout is implemented to explore binary image oversampling. Both the digital and analogue readout electronics were conceived and designed to be simple and practical.

A nanosecond electronic shutter is created by one of two time-gate pulse generators. In the original 'AA' chip revision, the two time-gate pulses are routed through balanced clock trees on both sides of the array into row-wise time-gate drivers. In the updated 'AB' IC revision, the right hand side clock trees and time gate drivers are removed. In both revisions, the driver circuit selects one of the two time-gate pulses to drive onto each row of the imaging array. Further row driver circuitry shown at the left of the diagram, handles the row select and read signal, the pixel reset, and the time-gate disable functionality. Both the X and Y (row and column) addressing decoders are binary to 'one hot' thermometer code converters. Each of these blocks is individually expanded upon in this chapter.

The image sensor was bonded into a Ceramic Pin Grid Array (CPGA) 144 pin package. A PCB was designed and fabricated to support this image sensor (and other sensors) which is shown in Appendix 3 ('PCB 1919A') along with the pin out from the image sensor.

The sensor is controlled by an FPGA, which handles exposure control, pixel addressing, readout timing, and oversampling and manages the complete data pipeline from sensor to computer. The Verilog HDL-based firmware, written to programme the FPGA, is described at a functional level in this chapter.

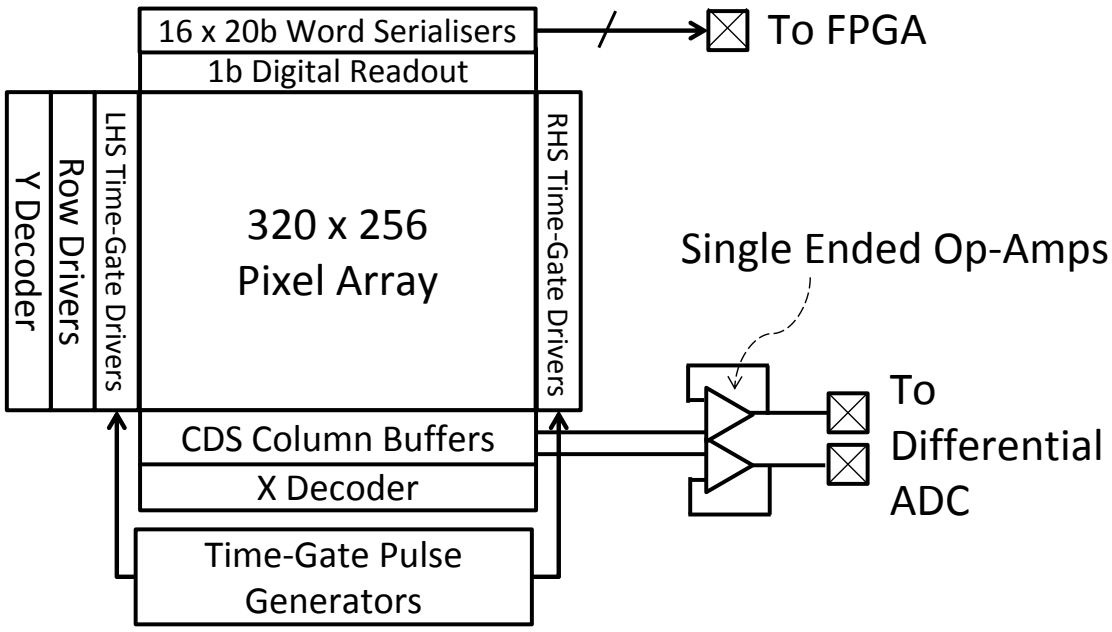


Figure 5.2. Overview block diagram of the SPC Image Sensor ('AA' IC revision).

5.2.1. Revised Pixel

The test array pixel described in Chapter 4 is a first step to realising a high spatial resolution single photon counting SPAD image sensor with sub-10μm pixel pitch. It is limited in fill factor (3.2%) because of use of the isolated diode. To achieve a significant increase in sensitivity, the use of double strip well sharing is proposed in Chapter 3 and implemented in this image sensor to increase this to 26.8% whilst lowering the pitch to 8μm. In doing so, the available space for pixel electronics is reduced. Therefore, the pixel circuit described in Chapter 4 is redesigned to be as compact (in physical dimensions) as possible whilst retaining the counting and time-gated functions. The SPAD-node pull-up and FPGA pulse-test transistors are removed leaving a pixel with nine NMOS transistors. Table 5.2 provides a description of each of the pixel signals and supplies. Figure 5.3 illustrates the pixel schematic with example voltage waveforms shown in red. As depicted in the figure, the pixel supply 'VRT' is vertically gridded to ensure only one pixel is active on the thin metal supply line

during a sequential readout, lessening the risk of a horizontal IR drop causing a FPN or horizontal shading[1]. The high voltage SPAD operating supply 'VHV' is gridded both horizontally and vertically and decoupled on-chip to its respective ground 'SPAD GND' by a large 0.7nF MOM capacitor. The other row driver supplies and the pixel 'VRT' supply have poly decoupling capacitors at the edge of the array.

Signal	Description	Source	Typical Range (V)
GND	Substrate and quiet analogue ground, bulk connection for all pixel transistors	Chip substrate and PCB analogue ground	0
VRT	Pixel supply voltage	PCB DAC-controlled regulator	2.7
VHV	SPAD operating high voltage	PCB 0V to 30V boost circuit and driver.	14.2 to 16.2
SPAD GND	SPAD operating voltage ground		0
VQ	SPAD passive transistor quench voltage	PCB DAC	0.6 to 1.5
VG	CTA gate bias voltage	PCB DAC	0 to 3.3
VS	CTA source bias voltage	PCB DAC	0.1 to 1.0
Time Gate	Electronic shutter	Time gate generator and row driver	3.3
Disable	Pixel input pull down	Disable signal row driver	3.3
RST	Pixel reset	Reset signal row driver	3.6
RD	Pixel read out row select	Read signal row driver	3.3

Table 5.2. Table of supplies, input and output signals of the revised 9T SPC pixel and descriptions of the respective sources.

Transistor	Description	Width (µm)	Length (µm)
M1	SPAD PQPR Transistor	0.3	0.97
M2	Time Gate Enable	0.3	0.35
M3	Time Gate Disable	0.3	0.35
M4	Counter and APS Reset	0.3	0.35
M5	Counter Source Follower	0.3	0.8
M6	Counter Drain	0.3	0.8
M7	APS Readout	0.3	0.35
M8	APS Read Select	0.3	0.35
MC	MOS Capacitor (Poly on Substrate)	0.58	3.7
CF	MOM Fringe Capacitor in Metal 1	0.12	6.0
	MOM Fringe Capacitor in Metal 2	1.08	3.3

Table 5.3. SPC imager pixel transistor and MOM capacitor sizes.

[1] In both SPCIMAGER_AA and SPCIMAGER_AB, the insertion was missed of the vertical metal grid via on every second row producing an alternating horizontal IR drop shading pattern which was compensated for by the CDS readout timing. This is described further in Chapter 6.

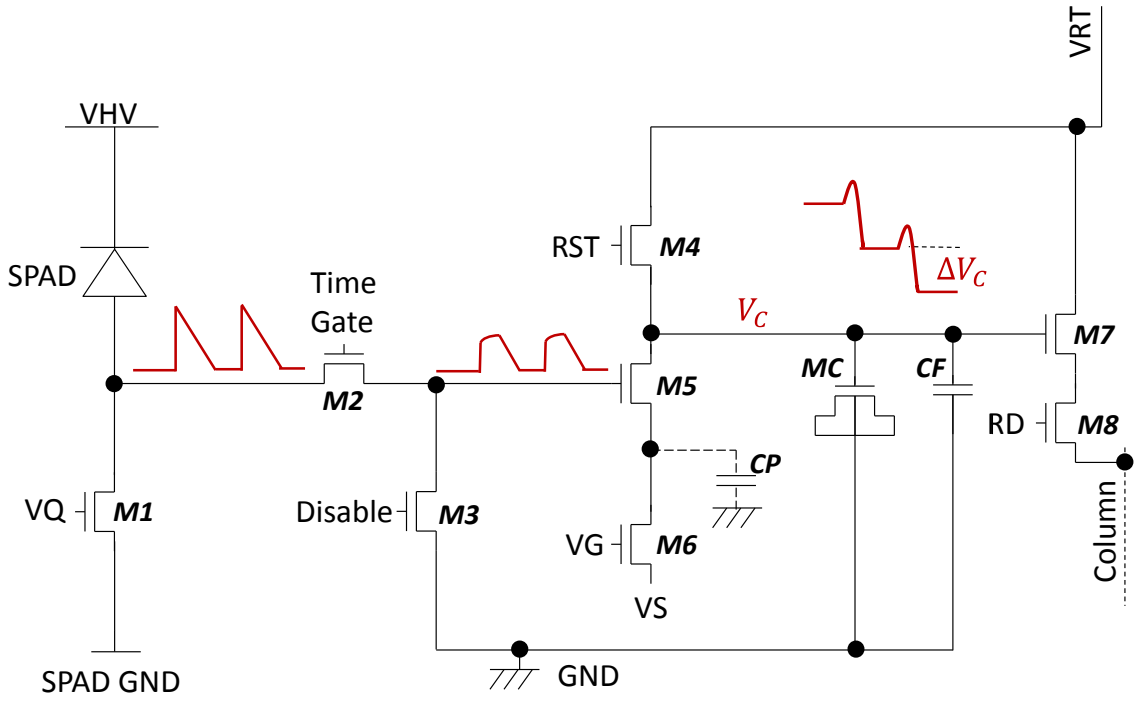


Figure 5.3. SPC SPAD-based image sensor pixel schematic. Nine NMOS-only transistors provide time-gated analogue SPC operation. Example voltage waveforms are shown in red.

The minimisation of every transistor size was evaluated to produce the most compact and efficient pixel layout. Table 5.3 lists the widths and lengths of each of the pixel NMOS transistors and the MOM fringe capacitors shown in Figure 5.3. All eight transistors and the MOS capacitor are implemented using the thick gate oxide transistors ('GO2' with 0.35µm feature size) available in the process designed for 3.3V operation. Time gate 'M2' and 'M3', APS reset 'M4' and readout source follower 'M7' and select 'M8' are all minimum sized. MOM capacitor 'CF' consists of both metal 1 and metal 2 fringe capacitors. A transistor shared active 'strip' layout is implemented, sharing drain and source connections to reduce the number of active region breaks between devices. Figure 5.4 illustrates the layout of the pixel transistors, demonstrating that only one break in the active region is needed. The area of metal 1 is maximised over transistors to optically shield the drain and source terminals to reduce photon-induced leakage. The MOS capacitor 'MC' (10fF approximately) is polysilicon over a grounded N+ active region. The pixel layout mirrors horizontally on the lines marked 'mirror' in Figure 5.4, such that the next pixel in the same row is a mirror image. The dark shaded area is the pixel in the row below, which is a 180 degree rotation of the unshaded pixel above. This three part layout strategy of shared active strip, mirroring in X direction, and rotation in the Y direction ensures the most compact layout of the pixel transistors possible.

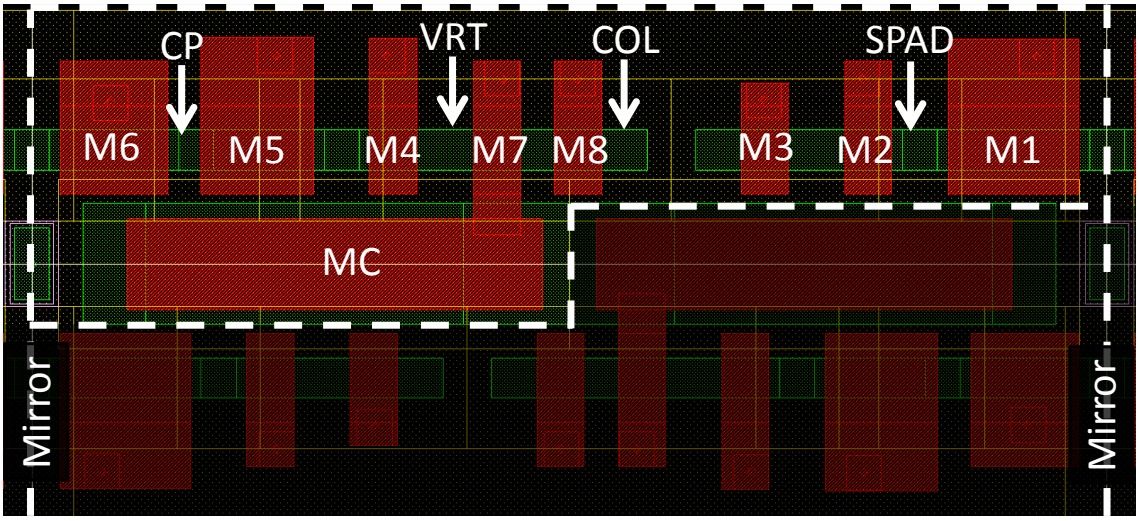


Figure 5.4. Pixel transistor layout demarcated by a white dotted line. Transistors and nodes of interest are shown as white overlaid text. Active silicon (non-STI region) is represented as green, poly silicon as red, light yellow on black is N+, and light pink on black is P+. A pixel from the next row is shown in the darker region. Mirroring lines indicate where the neighbouring flipped pixel layout meets.

Further to the pixel electronics, the SPAD layout is designed to obtain the highest fill factor in the confined 8µm pitch. As previously discussed, the double strip local well sharing layout approach offers the greatest fill factor attainable for a monolithic image sensor pixel to the detriment of different MTF's in the X-axis and the Y-axis. Also, a non-circular anode attains the highest possible fill factor whilst maintaining a square pixel. The layout of the whole pixel is shown in the following Figure 5.5, split into four parts (a) to (d) to illustrate it with no metal and with the subsequent addition of the three metal layers. The use of metal three ('M3') prohibits the use of the 'cavity' CAD layer, an oxide etch in ST's process which reduces the optical stack height of the oxide above the pixel for improved optical transmission into the photo active region. Figure 5.5(e) denotes the function of each line in metals 2 and 3 in the pixel array.

The mirroring and flipping of transistors causes non-matched current directions in the readout source follower and the CTA analogue counters (M5 and M6) which will contribute to FPN. However, without this layout technique such a small pixel layout could not be achieved.

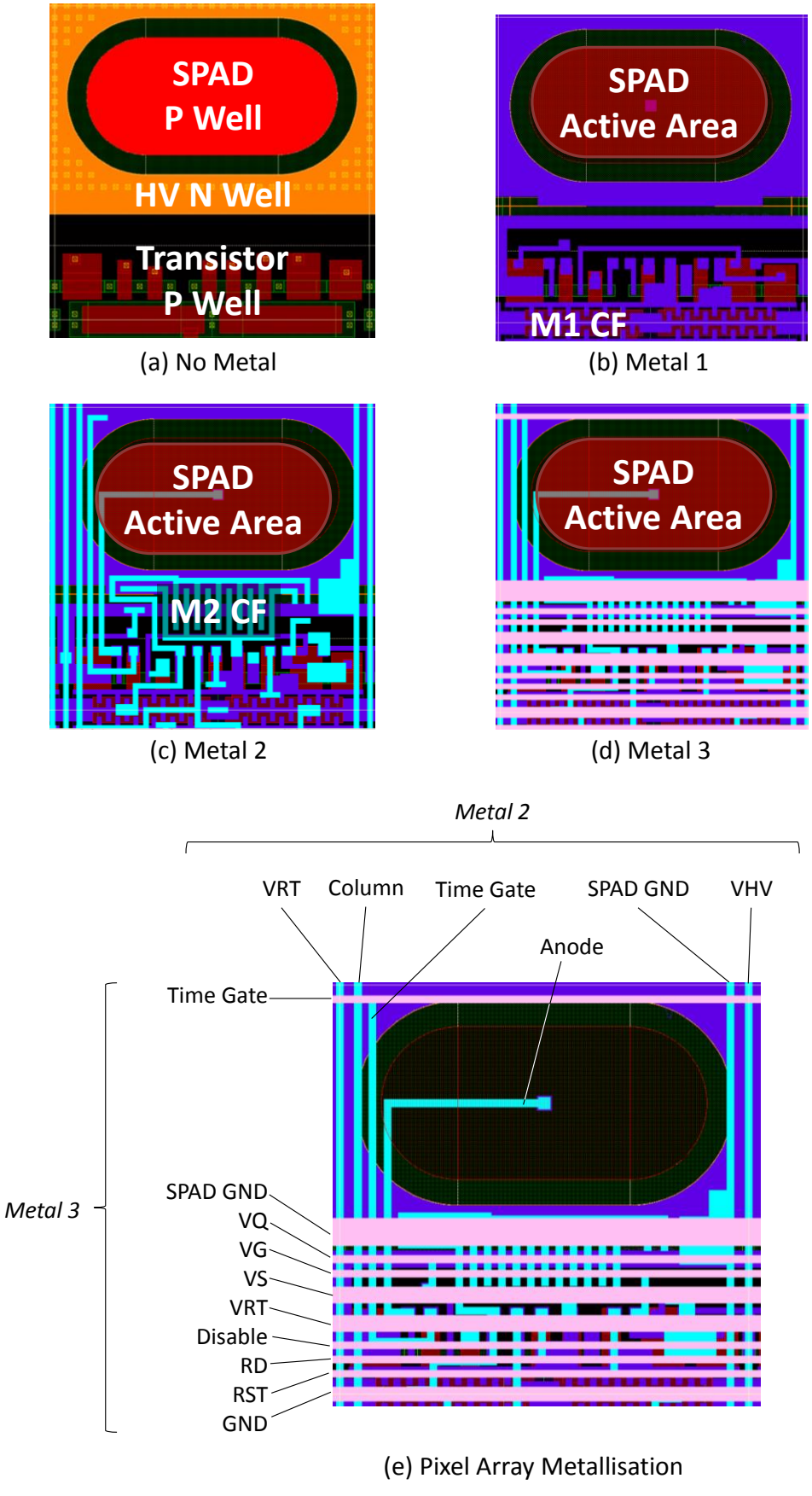


Figure 5.5. SPC image sensor pixel layout: the non-circular anode attains 26.8% fill factor in 8μm pixel pitch. Metal 1 and metal 2 MOM capacitors are denoted by 'M1 CF' and 'M2 CF'.

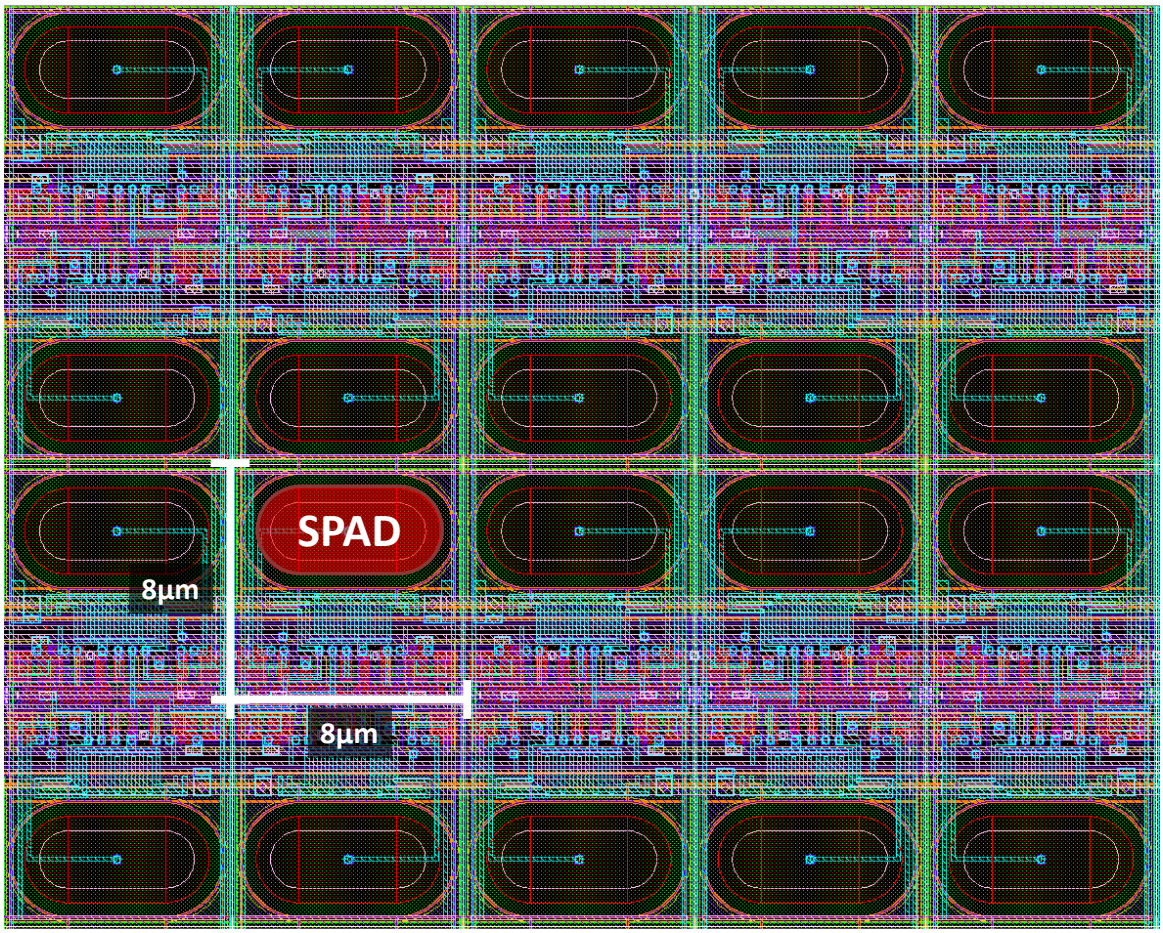


Figure 5.6. Image array layout: five rows and five columns are shown highlighting the double strip local well sharing approach giving high fill factor to the detriment of non-uniform spacing in the Y-dimension.

In Figure 5.5(e), the bias lines (VQ, VG) and the read and reset signals (RD, RST) are minimum width as no IR drop will be induced. The VS supply is double minimum width as it will have a current flow from the CTA discharge node. VQ, VG and VS are shielded from the high frequency pulses on the 'Time Gate' and 'Disable' row lines and to a lesser extent 'RD' and 'RST'. The 'Time Gate' row line is isolated from the rest of the row signals and supplies to avoid electrical cross-talk. The VRT line is gridded in both metal 2 and 3 as it will have a high current consumption during read and reset. The VHV supply is gridded in metal 1 with thin straps between rows in metal 2. The SPAD GND line is gridded in metal 2 and metal 3 and has triple minimum width in metal 3 to achieve a low IR drop during SPAD operation (which would create a dip in sensitivity in the centre of the array in high light). The layout of the pixel array is shown in Figure 5.6 with a five by four pixel example where the mirroring and rotation of pixels is visible.

5.2.2. Analogue Readout

The analogue readout chain is shown in the following Figure 5.7. Each column bus in the array is attached to a current source and correlated double sampling (CDS) sample and hold buffer performing a delta reset sampling operation [46]. Each column-wise CDS buffer is sequentially scanned out using a delta difference sampling (DDS) operation for vertical fixed pattern noise reduction using a crowbar switch [158]. The CDS buffers are connected on a differential analogue readout bus to two single-ended buffer-connected operational amplifiers (op-amps) driving out to an external *'Texas Instruments'* 14b differential input pipelined ADC ('TI ADC14L020') connected to a 3.3V regulated supply. Table 5.4 highlights a number of the ADC specifications reported from the datasheet.

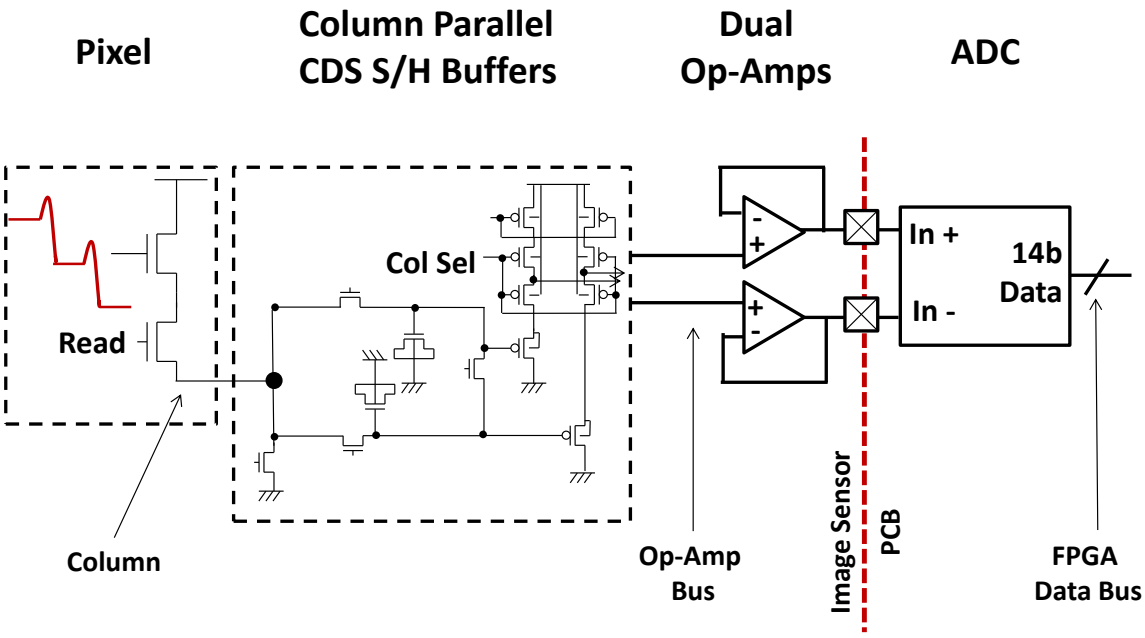


Figure 5.7. SPC image sensor CDS analogue readout chain. The boundary between the image sensor and the PCB is denoted as a red dashed line.

Parameter	Value
Power Consumption	150mW
DNL	±0.5 LSB
INL	±1.2 LSB
SNR (10MHz Input)	74 dB (typical)
Spurious Free Dynamic Range (SFDR) (10MHz Input)	93 dB (typical)
ENOB (10MHz Input)	11.7b
Input Capacitance (Clock Low)	4.5 pF
Input Capacitance (Clock High)	11 pF

Table 5.4. Texas Instruments ADC performance table.

5.2.2.1. Column Parallel CDS Buffer

The column parallel crowbar CDS buffer was designed and implemented in layout by Prof. Robert Henderson to fit in the 8µm column pitch. CDS is now a standard technique, in CISs, to cancel out the effect of output source follower threshold variations, kT/C noise and 1/f noise in an output image. It is achieved by capturing a 'dark' sample of the pixel value containing only the noise, and after a period capturing the signal (and noise) and subtracting the two. There are two variations of CDS timing: true CDS and delta reset sampling [46]. True CDS timing samples the reset signal before an exposure, and samples the signal level after. However, without two in-pixel storage mechanisms (to store the reset value and the signal value) it is inefficient and limited to only capturing an image sequentially line by line (the slowest rolling shutter technique). Delta reset sampling reverses the timing of the two operations: first, sampling the pixel signal and secondly sampling the subsequent reset. This has the advantage of both being able to capture a global shutter image and resetting the pixel before the next exposure. However the reset sample is taken from the next triggered reset, so the noise and offset subtractions are not truly correlated and cumulatively increase by square root of two [46]. Delta reset sampling timing is used in the column parallel CDS buffer in this work on the SPC image sensor.

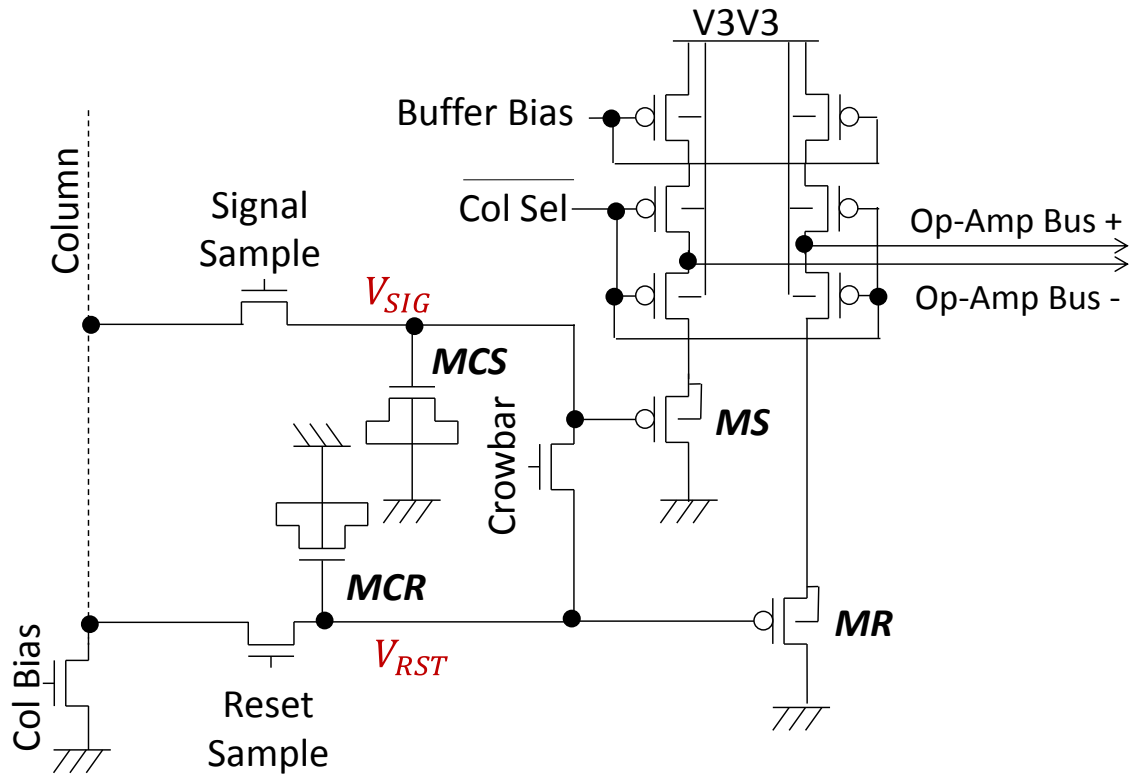


Figure 5.8. Column parallel CDS buffer with 'crowbar' VFPN removal switch for differential delta sampling (DDS) [158].

The sample and hold circuit diagram is shown in Figure 5.8. The PMOS source followers 'MS' and 'MR' are isolated in individual N-wells to reduce the body effect, a circuit technique to mitigate source follower non-linearity by keeping the gain closer to one over the whole input range. The crowbar switch implements a VFPN minimisation technique described in [158] under the banner of delta-difference sampling (DDS), where the ADC first takes a sample of the differential CDS voltage (V_{CDS} described in Equation 5.1), then the crowbar switch is pulsed on and off and the ADC takes a second sample of the shorted signal (V_{CB} in equation 5.2). A timing diagram is included later in this chapter in Figure 5.24. The two ADC codes are subtracted from each other removing the threshold variations in the PMOS source followers. This process is expressed in the following equations:

$$V_{CDS} = \alpha_{MR} V_{RST} - \alpha_{MS} V_{SIG} + v_{T_{MR}} - v_{T_{MS}} \quad \text{(Eq. 5.1)}$$

$$V_{CB} = (\alpha_{MR} - \alpha_{MS}) \frac{(V_{RST} - V_{SIG})}{2} + v_{T_{MR}} - v_{T_{MS}} \quad \text{(Eq. 5.2)}$$

$$V_{DDS} = V_{CDS} - V_{CB}$$

$$V_{DDS} = \alpha_{MR} V_{RST} - \alpha_{MS} V_{SIG} - \left((\alpha_{MR} - \alpha_{MS}) . \frac{(V_{RST} - V_{SIG})}{2} \right) \quad \text{(Eq. 5.3)}$$

$$V_{DDS} \approx V_{RST} - V_{SIG} \qquad | \alpha_{MR} = \alpha_{MS} = 1 \quad \text{(Eq. 5.4)}$$

where v_{TMS} and v_{TMR} are the threshold mismatches of each output channel, α_{MR} and α_{MS} are the cumulative gains of the MR and MS devices and the op-amp output channels. Equation 5.3 shows that the DDS technique supresses the offsets of each of the output channels. If the gains of the output channels are equal to 1, then as shown in Equation 5.4, V_{DDS} is equal to the ideal CDS voltage. However, if there is a slight gain mismatch between the channels then the output will have a signal dependent VFPN component expressed by the second bracketed term in Equation 5.3.

5.2.2.2. Output Operational Amplifiers

Two buffer-connected output operational amplifiers (op-amps) are included in the last on-chip stage of the analogue readout chain to drive the CDS buffer signals to the off-chip ADC at a relatively high rate. The single ended Miller compensated op-amp was designed, simulated and laid out by Luca Parmesan. In simulation, the op-amps deliver 70ns settling time driving an over specified 30pF load (for comparison the ADC input capacitance and PCB track capacitance is estimated at 12pF). The schematic of the op-amp is shown in Figure 5.9. The gain of the op-amps is unity by the buffer connection. The input range of the op-amps NMOS input pair is approximately 0.6V to 3.1V which matches the PMOS source follower output range from the CDS buffers which is approximately 1V to 3.1V. The push-pull output stage is Miller compensated.

The disadvantage of using two discrete op-amps to provide the differential path to the ADC is a global offset (chip to chip offset) will be introduced in the signal path.

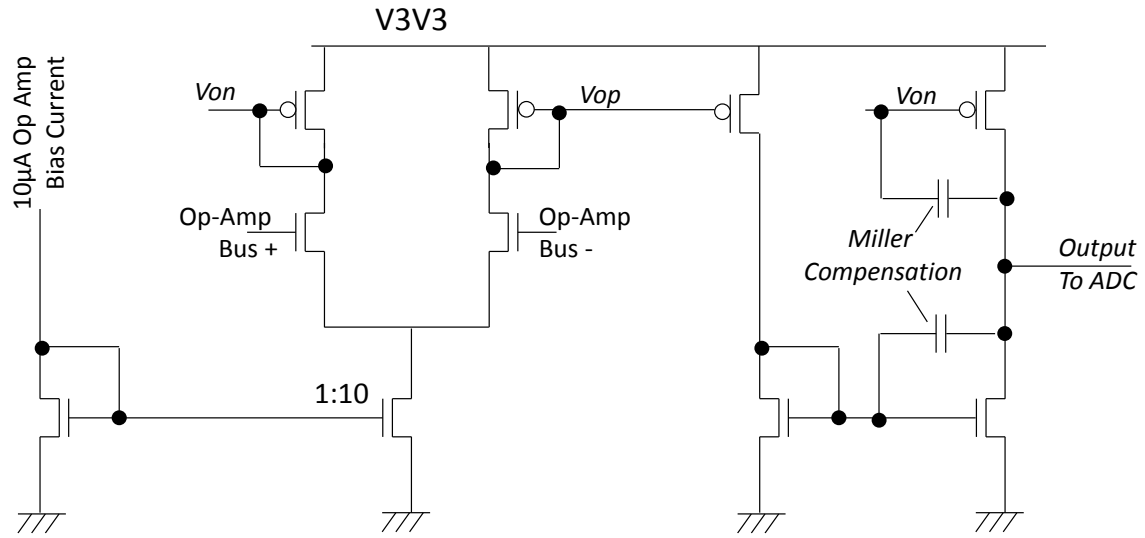


Figure 5.9. Schematic of output op-amp with NMOS differential pair and Miller-compensated push-pull output stage.

5.2.3. Digital Readout

The image sensor digital readout is designed to operate the device as a digitally oversampled binary image sensor. A diagram of the digital readout is shown in Figure 5.10. It is achieved by column parallel single bit latched comparators and sixteen parallel to serial converters each handling twenty columns. The simple readout is capable of serially streaming out the full QVGA array in 4800 clock cycles through the sixteen serial outputs. As an example, at a clock rate of ~25MHz, 5200 frames per second can be read out with a line time of approximately 800ns.

The coarse flash conversion and single bit digital readout is intended to function with the pixel biased in SCS mode with the highest counter step size. Each pixel output is either a high voltage if still in reset, or a low voltage if a single SPAD event has occurred and discharged the capacitor. This operates the pixel as a photon triggered dynamic memory. In this condition, the offsets and noise mechanisms such as kT/C noise, 1/f noise, source follower Vt variation, etc., are much lower in magnitude than the signal. Once these are input referred, these transistor-based noise sources become insignificant. Hence, the need for CDS is removed and the data conversion is performed in a single step.

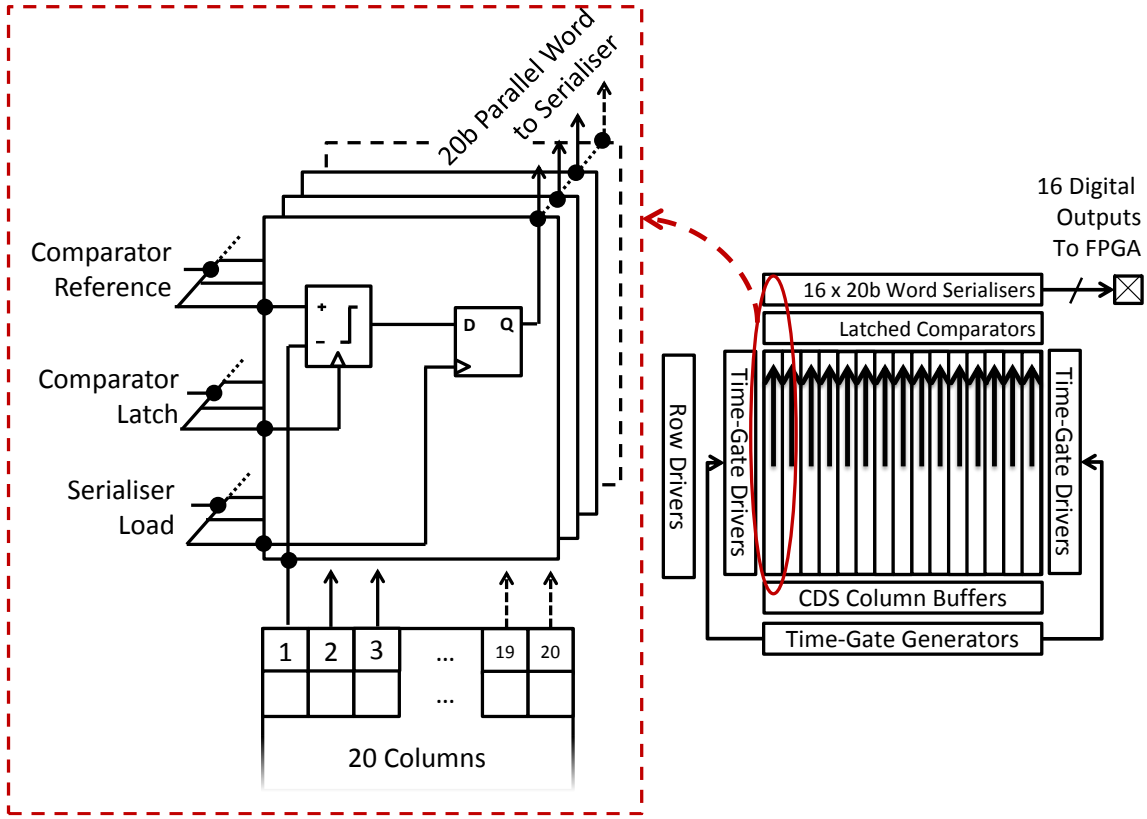


Figure 5.10. Digital readout: Each column connects to a latched comparator. Twenty comparator outputs are fed into a parallel to serial converter or serialiser. Sixteen serialisers read the full imaging array out as single bit data in 4800 clock cycles at thousands of binary frames per second.

Single sampling allows the row line time to be much shorter than conventional CIS line timing in the region of 100's of ns. With only two distinct voltage levels required, incomplete settling of the column lines is permissible, assuming the differential between the column and the reference is high enough to switch the comparator (avoiding metastable conversion with a long conversion time). The use of double sampling would substantially decrease the readout frequency, as the timing consists of two row settling periods (for both signal and post-reset sampling) with a reset period in between. This would be in the region of μs, which would at least half the maximum readout rate.

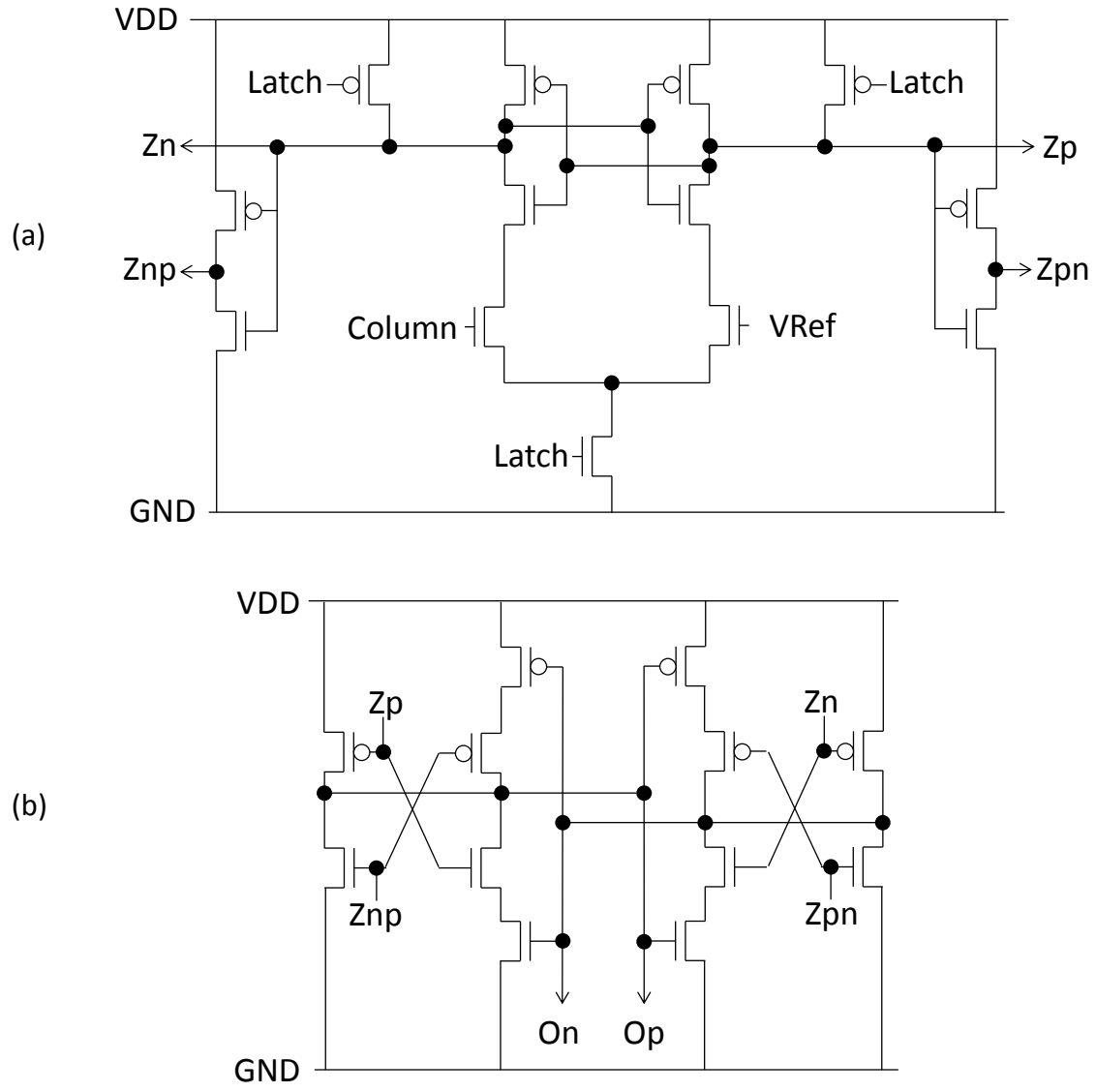


Figure 5.11. Fully differential dynamic latched comparator forming the front end of the column parallel single bit flash ADC of the digital readout [159]**. The upper diagram (a) illustrates the conventional latched comparator input stage and the lower diagram (b) shows the fully differential output stage.**

The schematic of the differential single bit dynamic latched comparator is shown in Figure 5.11 [159]. It has a conventional latched comparator input stage with a NMOS differential pair (Figure 5.11(a)). The column bus is connected directly to the positive side of the input differential pair. A global voltage reference is connected to the negative input of the comparator and is common to all 320 comparators, which is provided by an externally controllable off-chip DAC. The output stage in Figure 5.11(b) is fully differential with cross-coupled buffers, and a central cross-coupled inverter output stage.

Each column parallel comparator connects to a basic parallel to serial converter (or serialiser) for digital readout as illustrated in Figure 5.12(c). Only one side of each differential comparator output is connected to the parallel inputs of the serialiser via the mux input marked 'Column Data' in the Figures 5.12(a) and (b). Each serialiser accepts twenty parallel connections from the comparators and on assertion of a 'serial load' signal, samples the digital word for serial readout. Over the successive twenty clock cycles, the serial word is transferred off chip. There are sixteen of these serialiser blocks, each handling one sixteenth share of the imaging array as shown in the preceding Figure 5.10. The 'AA' chip revision parallel to serial cell is shown in Figure 5.12(a). To sample the first column data bit (column 19) the serial load must be low. This leads to a timing error, whilst sampling the serial data output off chip, if the load signal precedes the clock signal. Thus the improved Figure 5.12(b) parallel to serial cell was implemented in the 'AB' chip revision ensuring the output bit is only dependent on the serial clock.

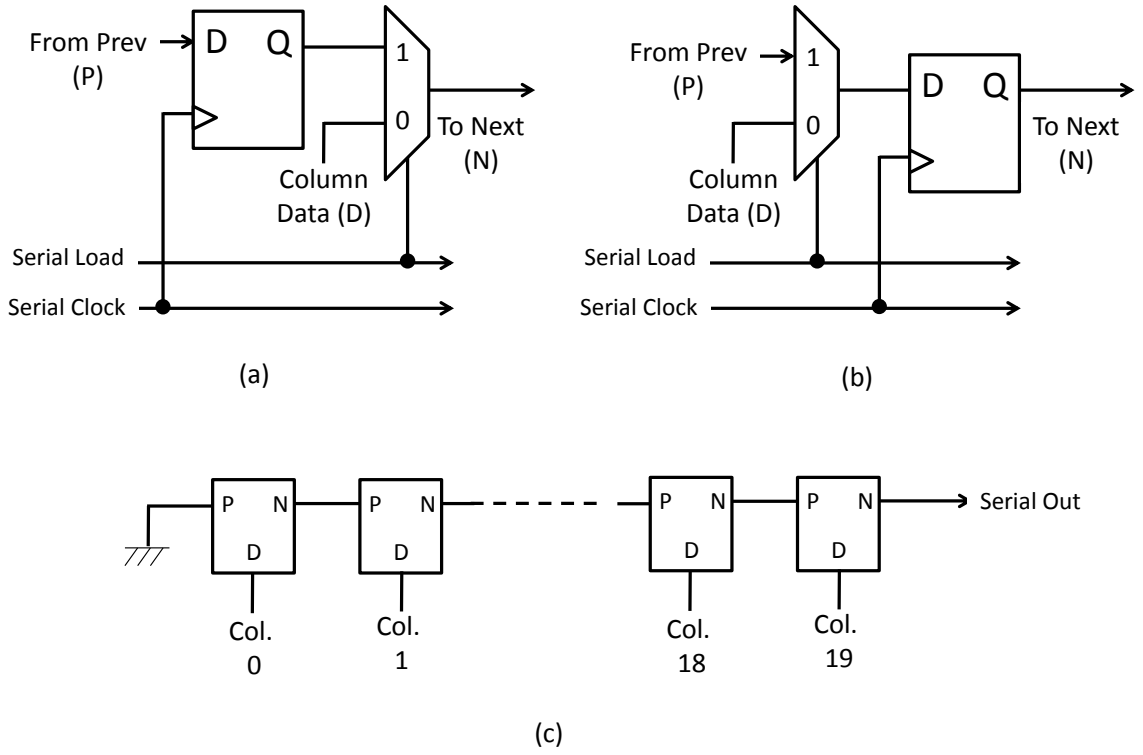


Figure 5.12. Sixteen 20b word parallel to serial converters are used for digital readout. (a) Original 'AA' revision parallel to serial cell. (b) 'AB' chip revision cell. (c) The serialiser chain made up of twenty parallel to serial cells.

The digital readout timing is shown in Figure 5.13. The minimum row readout time of the digital readout serialiser is the twenty clock cycles it takes to shift out the parallel 20b word. A row is selected by latching (with the address latch signal) the chosen row address into the X decoder, enabling that row's pixel read signal. The comparator is reset (pulling the comparator latch input low) whilst the column buses are settling. After a defined number of clock cycles the comparator is enabled and starts to make a decision, and eventually settles making a decision comparing the column input to the external voltage reference 'VRef'. There are two choices for the timing of the assertion of the comparator latch signal. For pixel signals with small swing, the column bus settles quickly and the comparator latch is asserted shortly after load. This allows the comparator to have a long settling time to account for the low differential voltage on the input (comparator settling is inversely proportional to the differential input voltage [159]). Alternatively for pixels with a large output swing, the column bus takes longer to settle and the column latch must be asserted closer to the end of the cycle. In this work a compromise is made, asserting the comparator latch mid-way (10 clock cycles) through the readout. At the end of the 20 clock cycles, the serialiser is loaded with the parallel data from the comparators and is serially streamed out whilst the next row is sampled by the comparators. Also the sensor row address inputs, from the FPGA, move on to the successive row in the array allowing the data bus to settle in the proceeding eighteen clock cycles and removing any possible timing errors in the row selection. The switching of the FPGA row address occurs whilst the comparator is still in reset reducing the effect of digital I/O switching noise on the comparator decision. This synchronisation of timing signals effectively pipelines the sequence (although not on a single clock cycle basis as per regular pipelining).

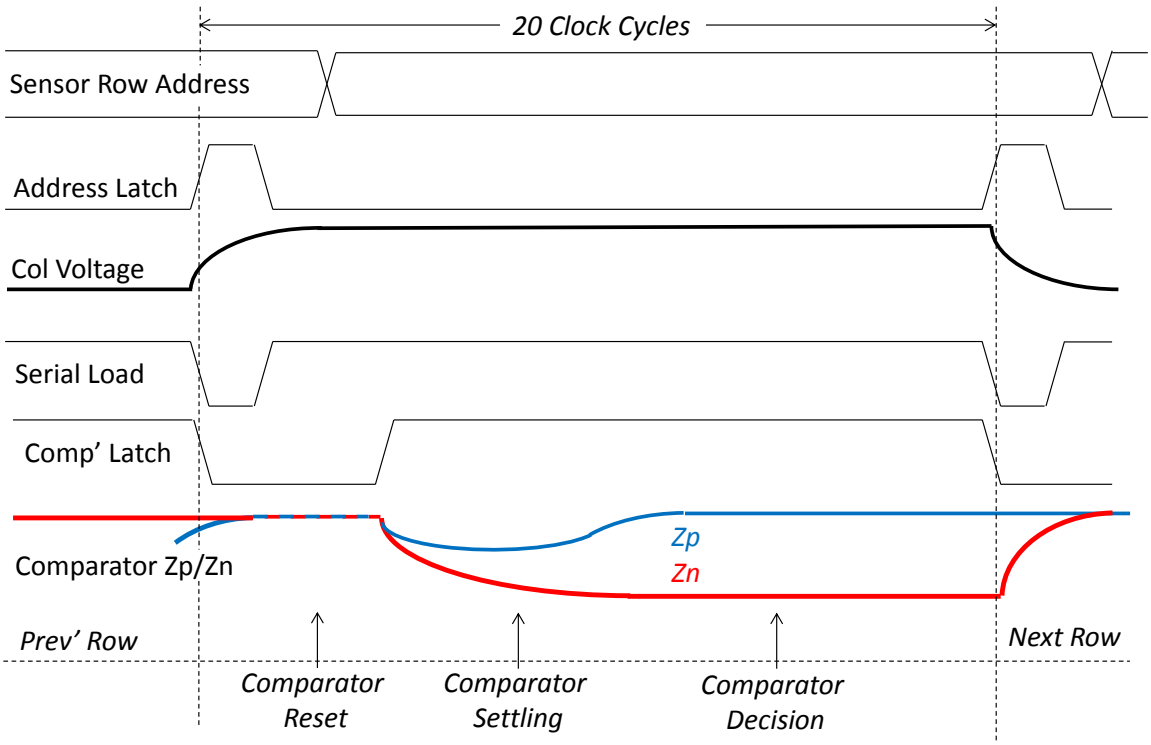


Figure 5.13. Digital readout timing. The minimum row time is the twenty clock cycles it takes for the serialiser to shift out its data.

5.2.4. Row Driver Logic and Address Decoders

This section describes the row driver circuits, the X and Y address decoders and the modes of image sensor operation. The remaining pixel signals are all biases or supplies which are supplied, generated and controlled off-chip and have on-chip decoupling.

There are four row-wise signals generated by row driver logic (the pixel signal name is shown in brackets):

- Row Read ('RD')
- Row Reset ('RST')
- Time Gate Enable ('Time Gate')
- Time Gate Disable ('Disable')

5.2.4.1. Row Drivers and Level Shifters

Each of the row driver circuits uses 1.2V CMOS (thin gate-oxide or 'GO1' devices) for the combinatorial logic and 3.3V tolerant CMOS (thick gate-oxide or 'GO2' devices) for the level-shifters and row drivers. Figure 5.14 illustrates the input logic, input level shifter and output driver common to all four row drivers. It shows the circuit diagram of the conventional level shifter circuit used. The read signal, enable and disable signals are supplied via the 3.3V supply (as shown in the figure) whereas the reset signal is supplied instead by a 3.6V supply.

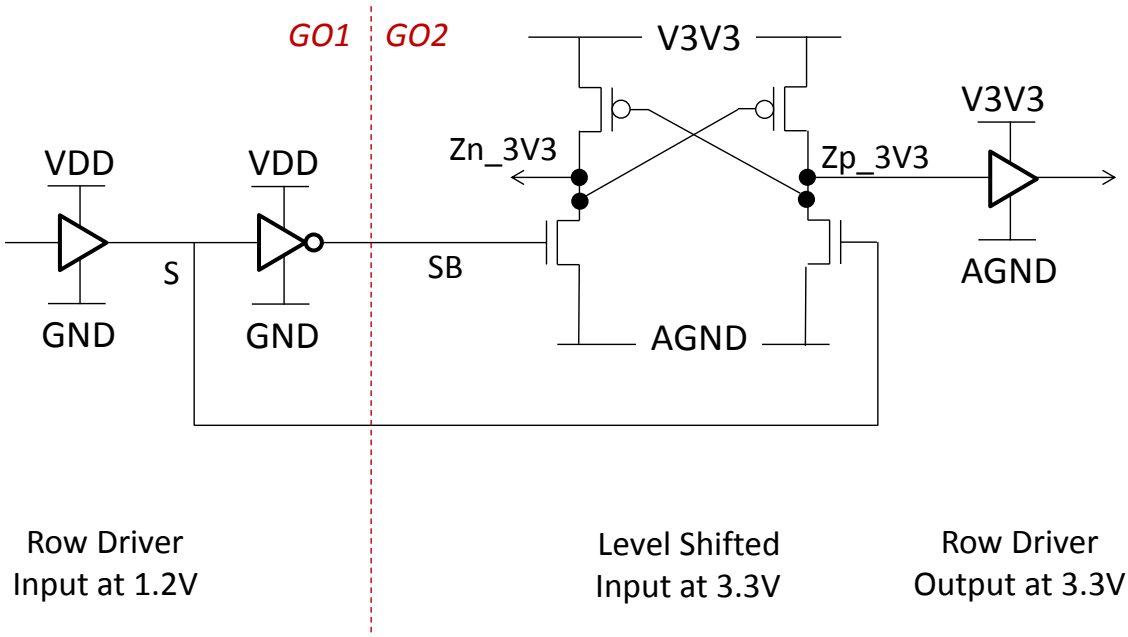


Figure 5.14. Row driver circuit diagram showing 1.2V GO1 input logic and 3.3V GO2 level shifter and row driver.

5.2.4.2. X and Y Decoder

Both the X and Y address decoders are 'GO1' binary to 'one hot' thermometer code converters. The 'Y-address' decoder generates the row read 'RD' signal (which also forms an enable signal for the other row logical operations). Figure 5.15 provides a schematic of the 8b Y decoder logic. The ROM-based block takes in an 8b row address from 0 to 255 (to cover all 256 rows) and latches the value using a load signal. Upon assertion of the load signal, the 8b word and an 8b bitwise-inverted word are fed into all 256 row decoders. Each row decoder has an individual input combination of either positive or negative bits from the two 8b words. The sixteen inputs are logically combined together though an AND gate guaranteeing that, for each binary word, only one row read signal becomes active. This same ROM-based binary to one-hot thermometer code decoder technique is used for the X decoder used for analogue readout to scan through the column CDS buffers. The X-address has a 9b word input to cover all combinations for the 320 columns.

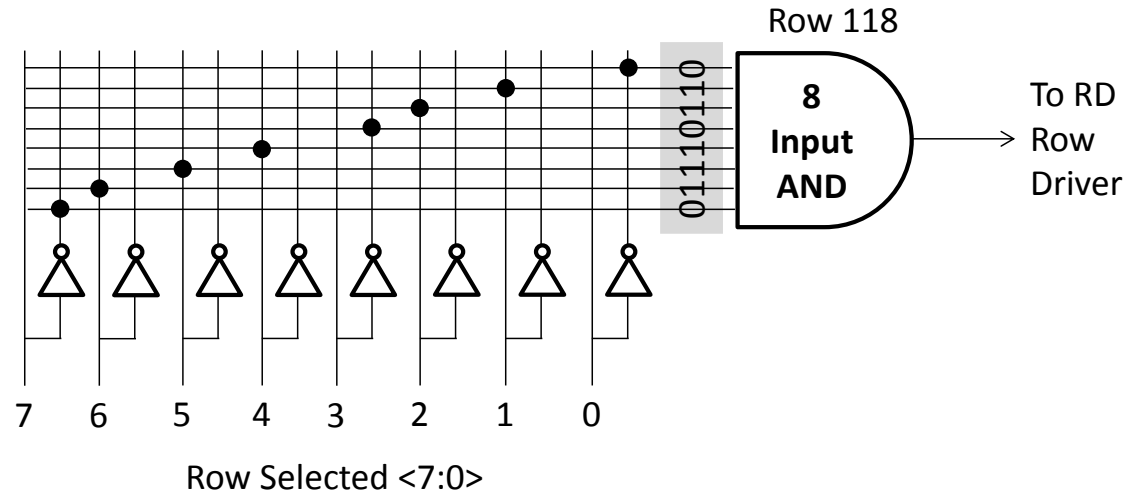


Figure 5.15. ROM-based binary to one-hot thermometer code decoder cell. The example shown is one decoder cell of the 8b Y decoder which will write the RD signal high for row 118.

5.2.4.3. Reset Logic

The 'GO1' reset logic is designed to produce an active high signal upon either of two input conditions. The first is a rolling reset activated by the row read signal logically AND'ed with the external reset signal from the FPGA timing control. The second is a global reset activated by the external reset AND'ed with a global shutter mode signal (again, both are input signals from the FPGA). Figure 5.15 illustrates these two reset logical functions OR'ed together to create the reset signal connected to the 3.6V level shifter and row driver block. The logic function is created with 4T NAND gates as they have less transistors than AND or OR functions and so are physically smaller to implement in the 8µm row pitch. The row level shifter drives the M4 pixel transistors in one row.

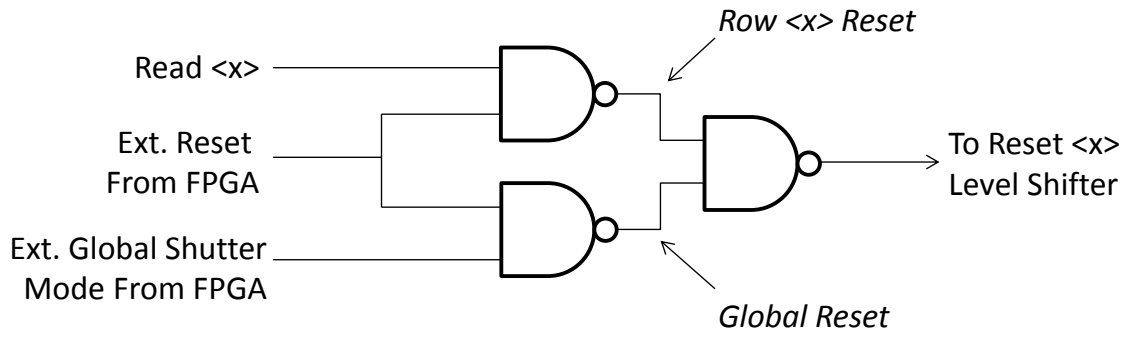


Figure 5.16. Reset row signal logical operation. The equivalent circuit of the parallel NAND gates is an AND function, and the single NAND performs an OR operation.

5.2.4.4. Pixel Time Gate Disable Logic

The 'GO1' disable row driver logic is designed to activate the pixel disable pull-down transistor M3, deactivating the pixel time gate. The disable is designed to operate under three separate conditions. The first is to activate the disable in a normal exposure when the external disable signal is pulsed and that row is not selected for readout. The second complements the first, activating the disable logic, whilst the row read signal is high. The last is a global disable function across the whole array upon the assertion of global shutter mode and the external disable signal (which can also be activated whilst the read signal is high for externally monitoring the pixel analogue output overriding the first condition, although this is not the primary purpose of this logic function). Figure 5.17 demonstrates the logic implemented to realise these three conditions for disabling the pixel time gate. The disable level shifter drives the M3 disable transistors in a single row.

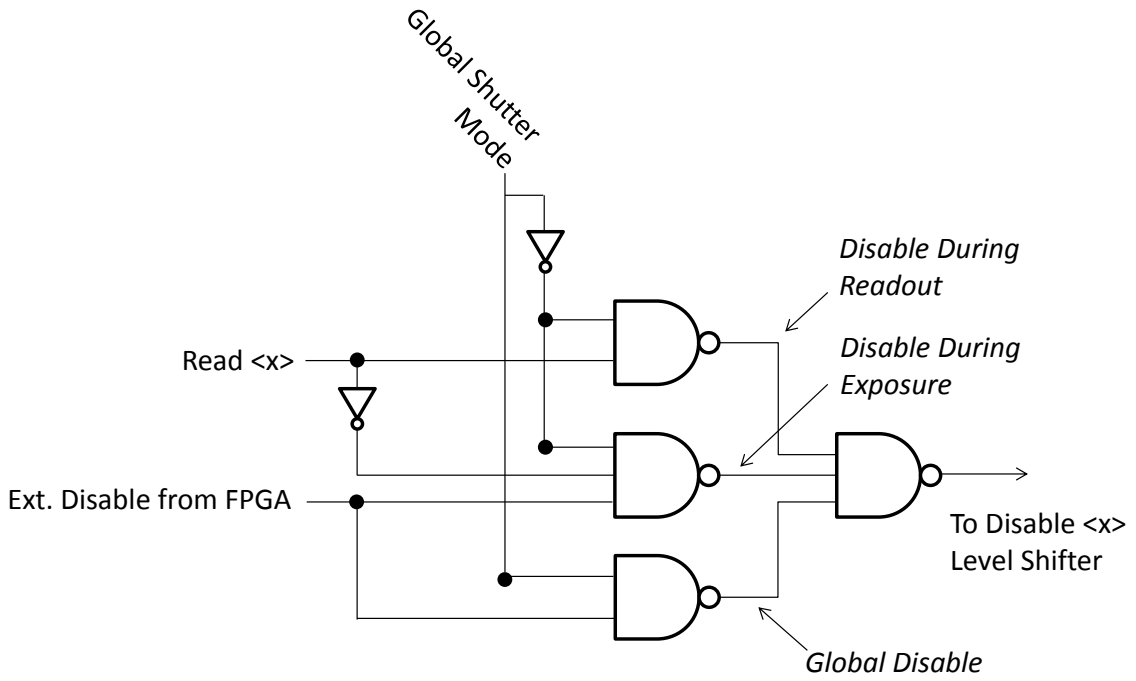


Figure 5.17. Time gate disable row signal logical operation. The equivalent circuit of the parallel NAND gates is an AND function, and the single three input NAND on the right hand side performs a three input OR operation.

5.2.4.5. Time Gate Driver Logic

There are two parts to the 'GO1' time gate driver logic: time gate input selector and activate or deactivate logic. The latter is shown in Figure 5.18(a), which ensures the time gate is not activated whilst a row is being read out unless the sensor is in global shutter mode. Activating the time gate whilst the read signal is high allows for externally monitoring of the pixel analogue output. Figure 5.18(b) illustrates the routing logic where a mux selects which of the two time gates signals 'A' or 'B' are routed into the pixel time gate M2 transistor. The mux select input is driven by a toggle flip-flop that resets low. The flip-flop toggles on assertion of either a global toggle switch (whilst in global shutter mode) or the read signal, either of which activates the toggle selecting the next time gate after each successive read.

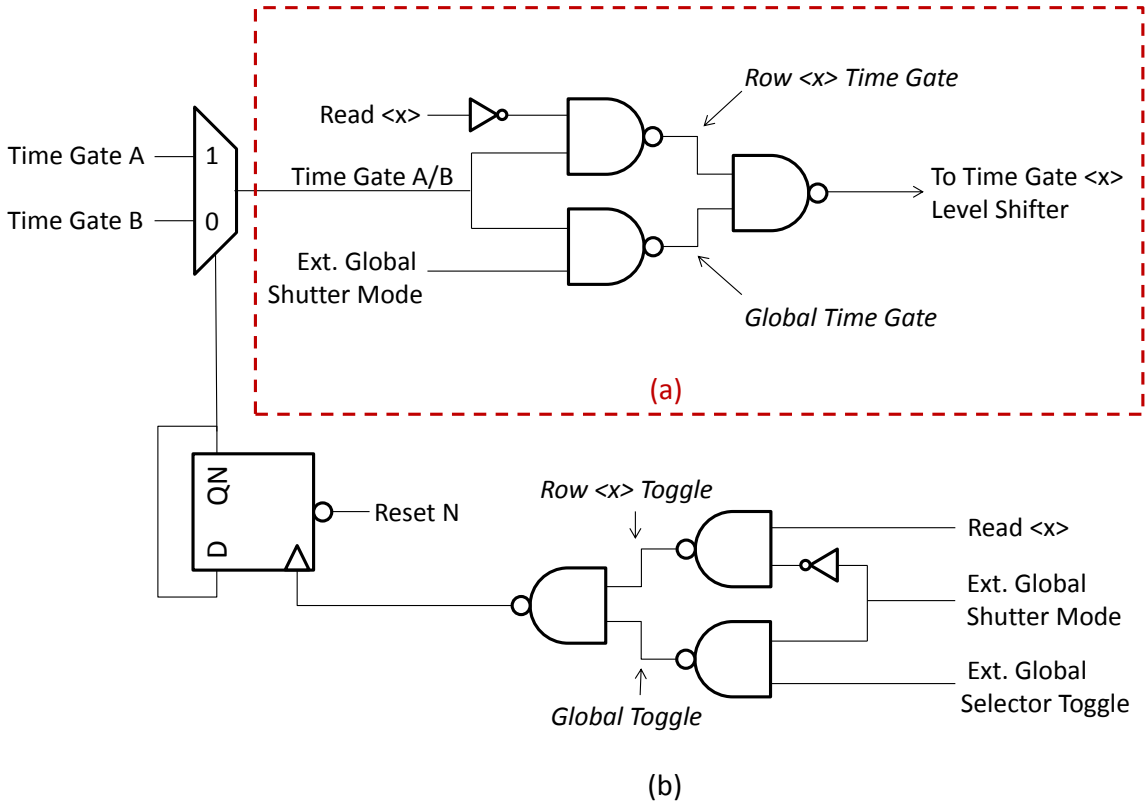


Figure 5.18. Pixel time gate enable logic: (a) activate/deactivate logic in the red box (b) the remaining logic is the time gate input selection between time gates 'A' or 'B'.

5.2.5. Time Gate Pulse Generation and Clock Tree

The pixel time gate is an electronic shutter for the image sensor and this section details the different shuttering modes and time gate operation. Two pulse generators on the sensor create the two time gates ('A' and 'B') required for two bin time-gated single photon counting. Figure 5.19 provides an overview of one of these time gate control and generation blocks. The pulse generator creates time gate signals with programmable timing offset and pulse width in 500ps steps. These pulse generators have two inputs, one from the FPGA and the other from a novel laser NIM interface circuit. The pulse generator can also be bypassed to allow the FPGA to fully control the pixel shuttering to take images with exposures longer than the limit of the pulse generator.

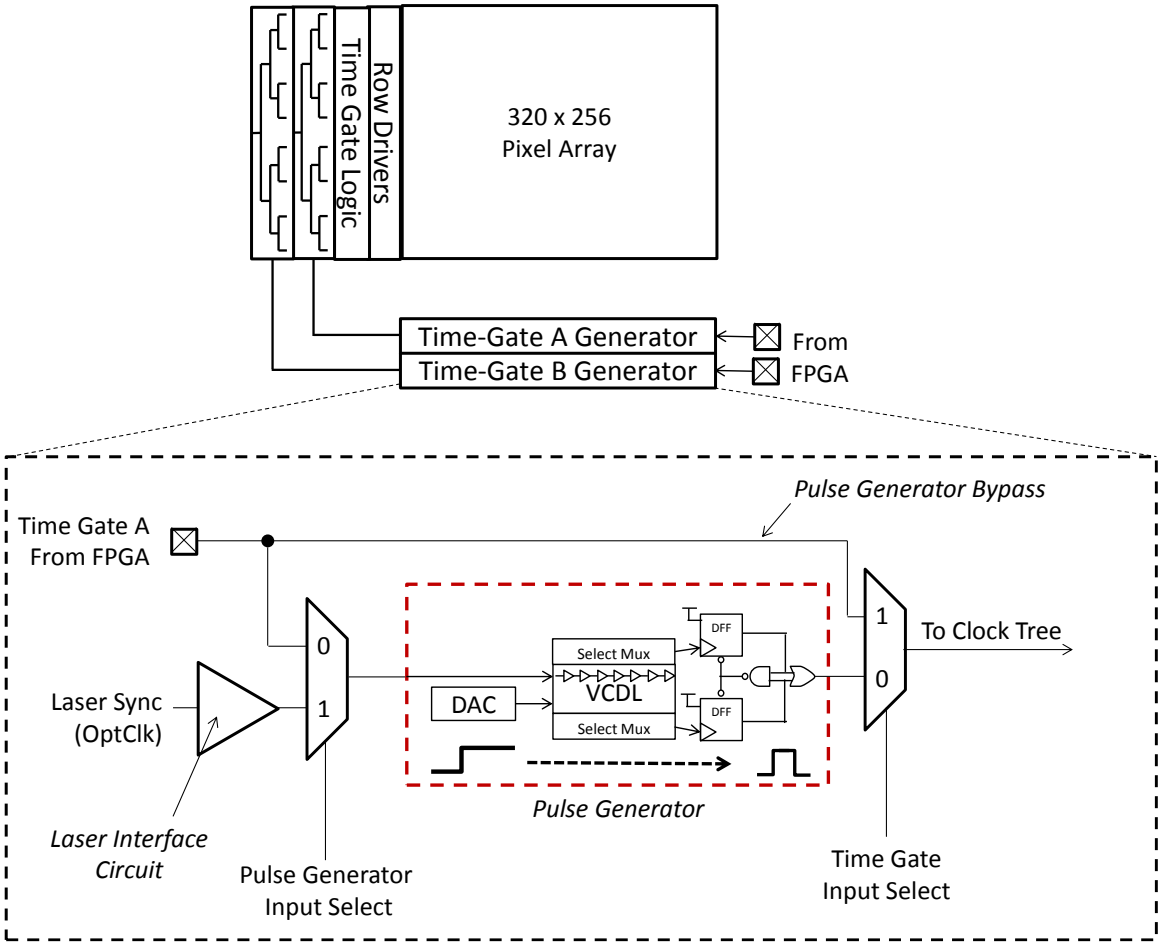


Figure 5.19. One of two time gate control and generation blocks.

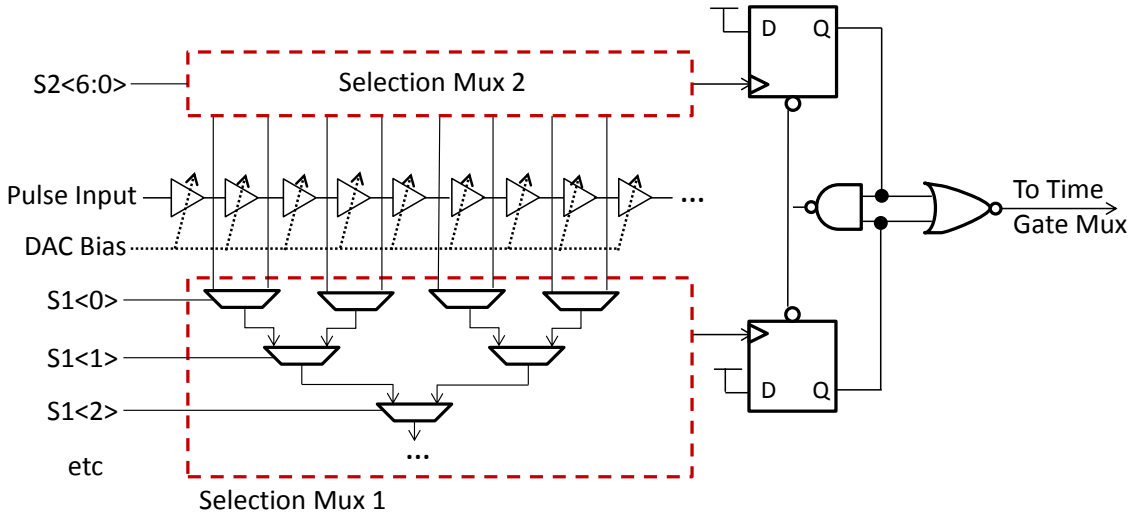


Figure 5.20. The pulse generator block comprising 127 delay element voltage controlled delay-line, two delay-line output selection muxes, and bang-bang pulse generation block.

Each time-gate signal is routed from the time-gate block at a central point at the bottom of the imaging array and distributed through a balanced metal clock H-tree. In IC revision 'AA', there are clock trees and time gate row drivers either side of the array (each driving half of one row) to half the total routing delay of the row signal across the array. However, in IC revision 'AB' the right hand side metal clock tree and row drivers are removed and the row lines were connected across the whole array due to a left to right mismatch in TOF images, increasing the delay across the array but removing the half array mismatch. The characterisation of the timing mismatch in both revisions 'AA' and 'AB' is detailed in Chapter 6.

The input to the pulse generator is chosen between the FPGA input and a laser synchronisation circuit. This laser circuit is designed to accept the negative going -600mV to -800mV peak 'NIM' (nuclear instrumentation module) pulse from a range of lasers in order to generate a time-gate pulse. It is a two-stage operational amplifier design with a resistor to purposely bias the amplifier with a voltage offset and time delay. An offset selection switch chooses whether the amplifier is sensitive to positive going edges of the input signal or negative going edges.

Each pulse generator, illustrated in Figure 5.20, is a digital to time converter (DTC) consisting of a voltage controlled delay line with 127 delay cells, and two delay-line output selection muxes. The voltage control is derived from a 4b current DAC with a non-linear diode load. The non-linear voltage generation allows the pulse generator to have finely controlled short pulses or long-pulses selected by the 4b DAC input word. Each delay cell has two current starved inverters, the second is loaded by an inverter feeding the output selection muxes and the first has a dummy matched load to balance both rising and falling edges down the delay line. The 127-input selection muxes are both seven-stage tree arrangements based on a single two input mux stage. The 7b binary selection input does not require a binary to thermometer encoder. The pulse generation block used is a conventional bang-bang circuit from a PLL phase detector [160]. The circuit takes the output of the two selection muxes, the first rising edge from either of the two muxes switches the pulse generation block high and the second rising edge switches the output low again as shown in Figure 5.21.

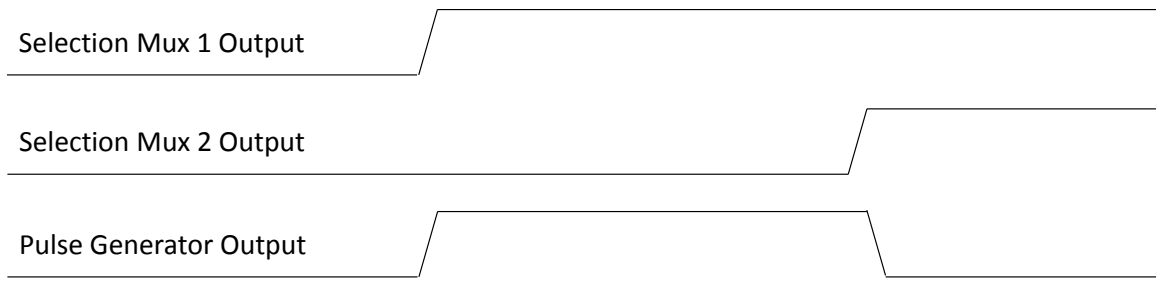


Figure 5.21. Pulse generator simple timing diagram.

5.3. Image Sensor Modes

The previous sections detailed the readout circuitry and control logic. This section will describe the four main modes of the image sensor and their respective timings. Table 5.5 summarises the four modes of image capture in terms of type of readout and the type of image formed. The FPGA Verilog HDL firmware written to handle the sensor timing and control, ADC data capture and data handling is described in an overview in the following section and as part of the sensor mode descriptions. Matching control software was written in MATLAB.

		Readout	
		Analogue	*Digital*
Type of Image	*Single Photon Intensity*	Single Photon Sensitivity Global Shutter Image	Digitally Oversampled Binary Image
	Time of Flight Depth Map	ITOF Image	Digitally Oversampled Binary ITOF Image

Table 5.5. The four image capture modes of the time-gated single photon counting image sensor.

5.3.1. Firmware Overview

Figure 5.22 presents a block diagram of the main blocks in the FPGA firmware, each of which is described in this section. The FPGA is an off-the-shelf Xilinx Spartan 3 on an Opal Kelly plug-in board. The interface between the USB and the FPGA is handled by an Opal Kelly communication block. An interface wrapper (highlighted in the 'OKConfig' box) was automatically generated by the generation script described in Appendix 3 ('OKConfig' script). The sensor has a basic three pin (clock, load, and data) serial interface for writing configuration data to the two pulse generators and for enabling the debug outputs. This is programmed by the 'Control Serial Interface' serialiser block which captures the state of configuration registers from the Opal Kelly interface and serially streams it out to the sensor.

The memory controller block comprises sixteen 75kb block RAMs, each 16b wide with 4800 addresses. The sixteen memories are written in parallel during the digital readout and accumulation, and sequentially during analogue readout. Once the sensor readout is complete, the memory controller streams the contents of the sixteen block RAMs to a 32kb FIFO, ready for USB data transfer. As the order of pixel outputs in the analogue and digital readouts operations is different, a round-robin procedure for the analogue data writes 20 addresses of one block RAM before moving to the next. This has the effect of maintaining the same data order for USB data transfer regardless of the image capture method. The memory controller and data handling is described in greater detail in the global shutter and digital oversampling sections.

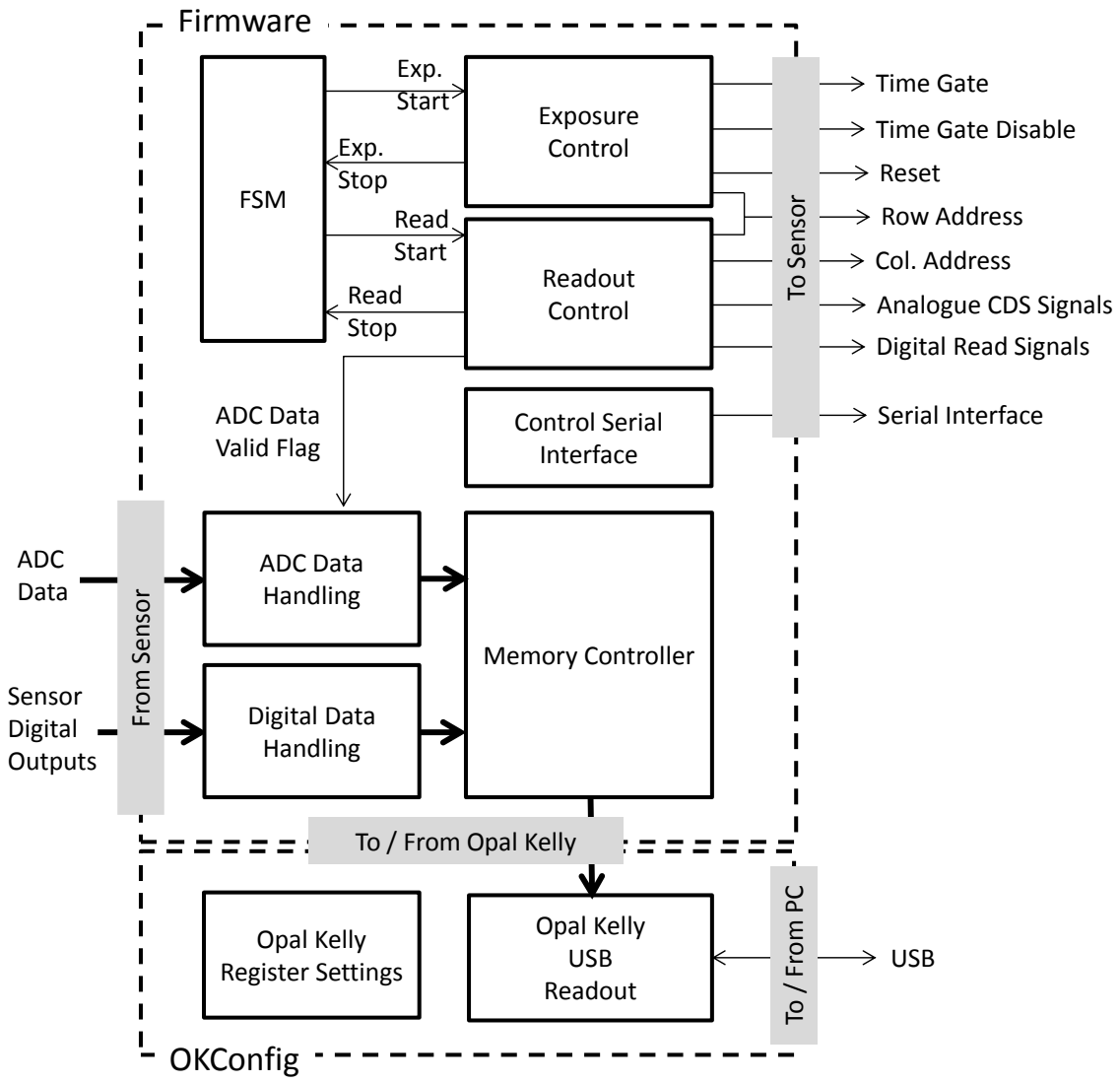


Figure 5.22. Block diagram of the FPGA firmware performing sensor timing, control and data handling. The section marked 'OKConfig' is an Opal Kelly interface wrapper detailed in Appendix 3.

5.3.2. High Sensitivity Global Shutter

An exposure is externally asserted by the user triggering the firmware's FSM to activate the 'Exposure Control' block. A 32b exposure register and counter is implemented in the exposure control block to handle a very wide range of exposure times. The system clock frequency used for operation defines the exposure time limits. The clock rate is tuned to match all three requirements (Chapter 6): op-amp settling, ADC operating frequency and digital oversampling operation. A typical 20MHz system clock rate offers a 50ns exposure step, hence the 32b range covers a minimum single clock cycle exposure of 50ns to a maximum exposure of 3 minutes 34 seconds. (The firmware also handles a setting of zero producing no exposure with the pixel held in reset.) The clock frequency is easily modified by a user, if required for an application or experiment, therefore the software control must handle the conversion correctly from the requested absolute exposure

time to the relative register setting. (This is handled in the MATLAB code for the sensor under the 'SetExposureTime' function).

Figure 5.23 illustrates the timing of capturing a global shutter image. Firstly, a reset is activated either as a rolling reset or a whole array global reset. The durations of both the rolling or global reset pulses are set by adjustable registers. Figure 5.23(a) displays the overall array timing of a rolling reset, global shutter exposure with rolling analogue readout. After the initial reset period, the exposure runs for the desired time by asserting the pixel time gate. This allows the pixel counter to integrate the number of SPAD avalanche events during the period of the global time gate. As illustrated in Figure 5.23(b), the disable signal operates as a non-overlapping signal with the time gate and so is set high during the initial reset and for the subsequent readout. Whilst the time gate is low, and the time gate disable is high, the pixel analogue counter is isolated from the SPAD. After the exposure, the 'exposure stop' signal is asserted and the 'readout start' signal is triggered by the FSM. As indicated in the diagram, the rolling readout of a full frame is a lengthy procedure compared to the initial reset; as an example, 1 microsecond initial reset per row produces a full frame reset in 0.24 milliseconds, whereas 5 microseconds per column readout requires 384 milliseconds per frame, and a maximum frame rate of 2.6 FPS. The initial rolling reset period may be adjusted to match the rolling readout period, to give all pixels the same time between reset and readout (equalising the leakage across pixels), at the expense of frame rate.

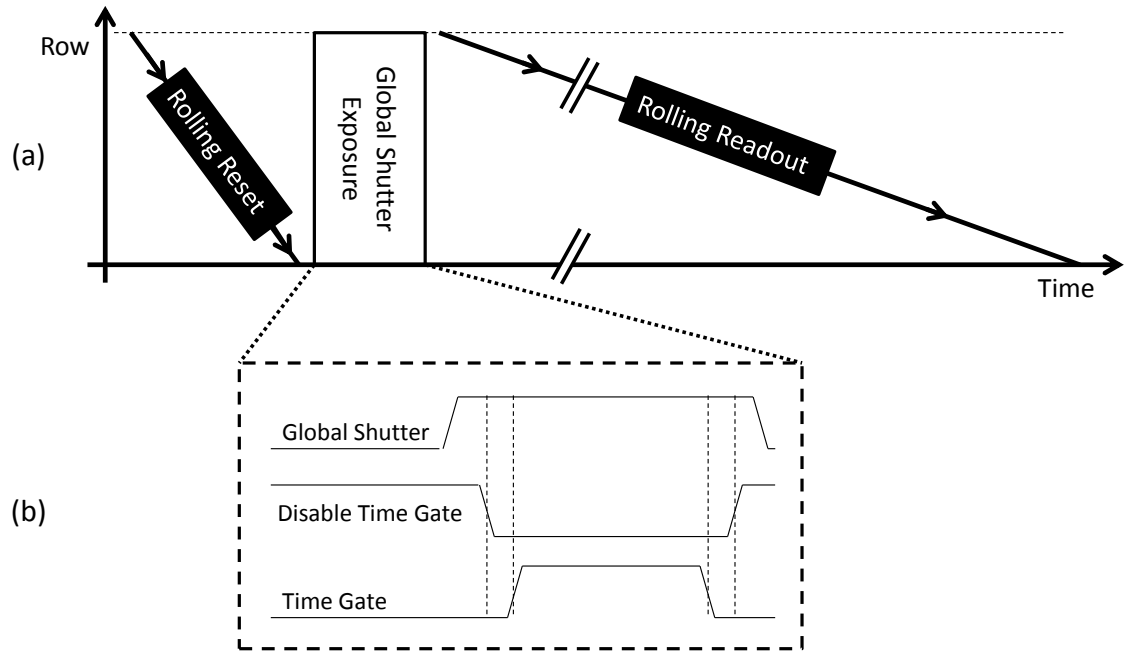


Figure 5.23 (a) Array timing diagram of global shutter image capture with rolling initial reset and rolling readout through the single channel analogue readout. (b) Highlighted global shutter exposure timing in the inset zoom-in figure.

The CDS delta reset timing proposed by Henderson et al. in [161] to reduce APS pixel-to-pixel FPN (PPFPN) and sampling capacitor crosstalk is implemented in the 'Readout Control' block. The first row is activated by the Y decoder, enabling all the read select transistors on that row. Figure 5.24(a) shows both CDS signals 'CDS Blk' and 'CDS Sig' are pulsed sampling the signal level onto the CDS buffer sampling capacitors. The reset signal is then held high, again for at least 2µs. One clock cycle after reset falls, the black sample 'CDS Blk' is taken. The column buffers have now sampled the first row and the column scan-out proceeds.

After the column parallel capture of signal and reset levels, the analogue readout timing is asserted sequentially by scanning each column CDS buffer out to the external ADC. Figure 5.24(b) illustrates how a column is activated by the X-decoder and the ADC output bus is allowed to settle before the ADC takes the first sample. It is assumed that the output bus is slew-rate limited as shown in the diagram. The ADC is left continuously running and the 'Readout Control' block activates the ADC data valid flag for one to four clock cycles instructing the 'ADC Data Handling' block to capture one or four samples which are then averaged and held in a register. The DDS crowbar timing proposed by Mendis et al. in [158], [162] is implemented in the 'Readout Control' and 'ADC Data Handling' blocks. The crowbar signal is pulsed and the column output bus is again allowed to settle before a further one to four samples are flagged as valid, averaged and held in a register. As described earlier in this chapter, this second crowbar ADC sample represents the digitised column buffer and op-amp offsets which are then subtracted from the signal. The output data word is then flagged as valid and passed into the memory controller and stored in block RAM. Once all the rows and columns have been read out, the ready signal is activated by the memory controller instructing the USB controller that there is a full image frame ready to be streamed out from the FPGA.

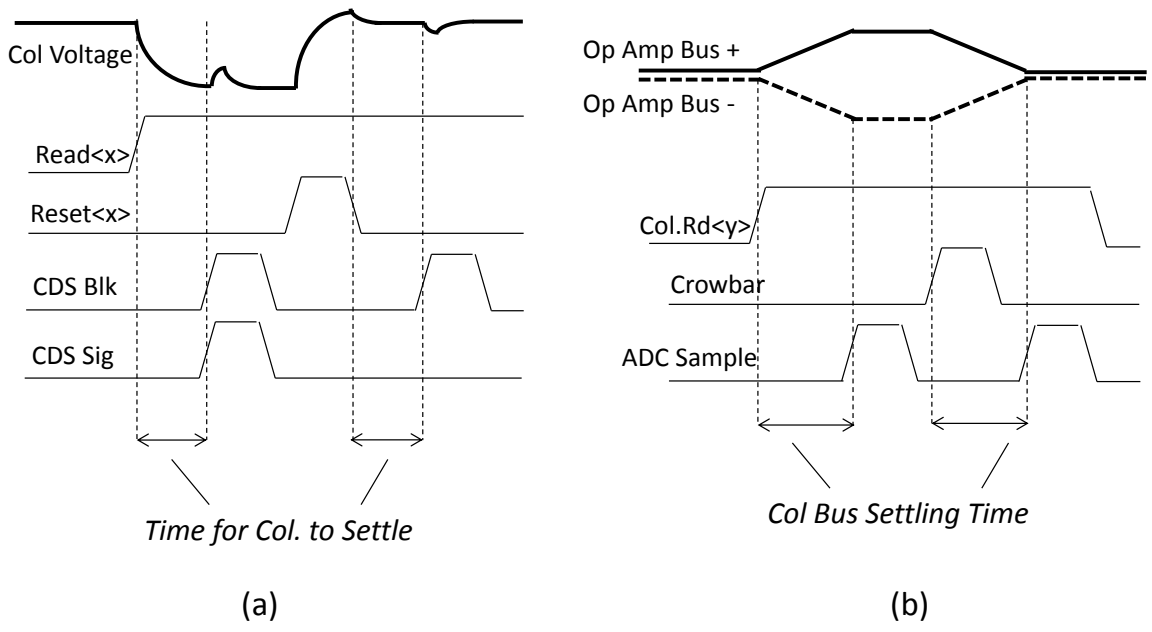


Figure 5.24. Timing diagram of the single channel analogue readout for one column. (a) CDS delta reset sampling mitigating PPFPN and column crosstalk [161]**. (b) DDS crowbar timing** [158], [162]**.**

5.3.3. Digital Oversampling

The combination of the pixel's SCS mode, the fast digital readout and the digital summation and memory capabilities of an FPGA, realises a SPAD-based oversampled binary image sensor which offers a first evaluation of the capabilities and properties of Fossum's QIS concept. The pixel counter is biased in the highest gain or current setting, essentially providing a SPAD-triggered dynamic 1b memory. Each single bit binary image from the sensor is referred to as a field image. The field image is serially streamed out across the 16 serial outputs in 4800 clock cycles. Using a 48MHz clock, a binary image field is transferred in 100µs at 10,000 FPS. Figure 5.25(a) demonstrates the digital mode array timing diagram with the capture of four fields and Figure 5.25(b) of a single field image.

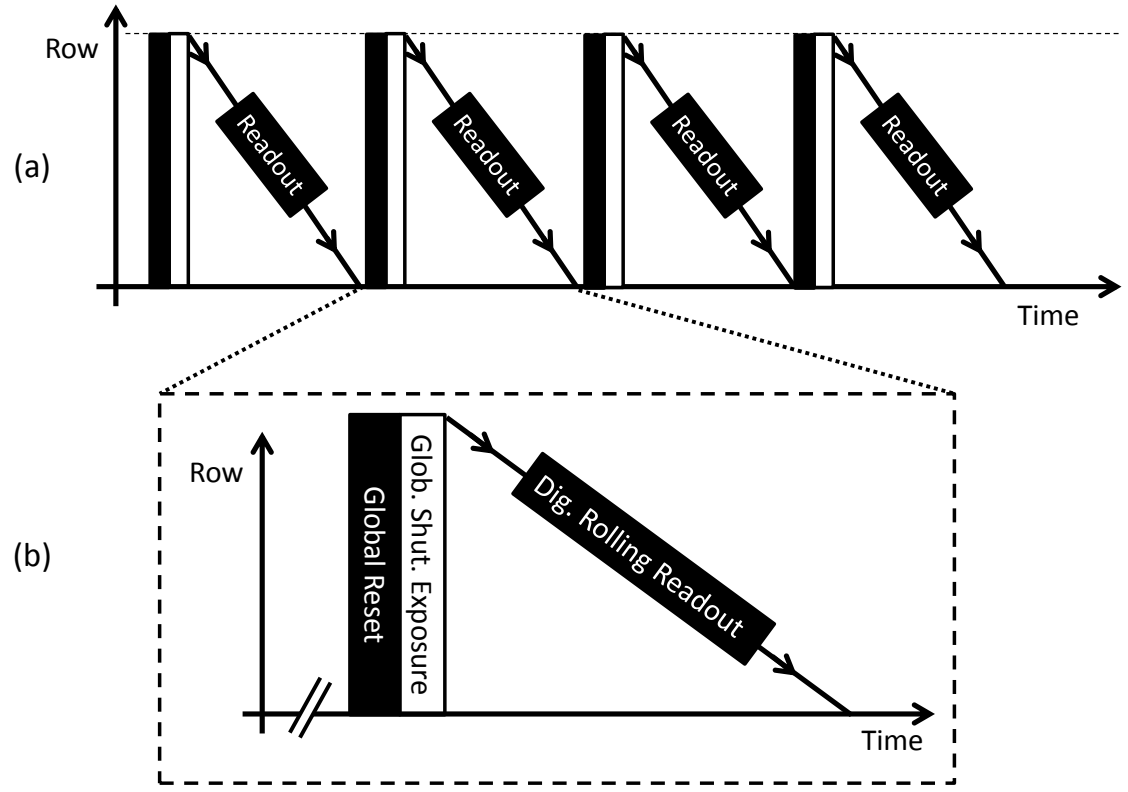


Figure 5.25 (a) Timing diagram of digital oversampled image frame capture with four short sub-exposures or fields. (b) The timing of a capture of a field image.

Fossum's digital film sensor proposal paper [45] discusses either spatial or temporal summation or both; this work explores temporal oversampling only. Field images are summed in the time-domain by the FPGA firmware taking a pixel's current value and summing it with the aggregated value from memory. For each doubling of oversampled field images, the output frame rate is halved from the FPGA. For example, forming an 8b image would require 256 field images, and the output frame rate from the FPGA would be a maximum of approximately 39 FPS if the sensor field rate was 10k FPS. Parallelism in digital design allows many concurrent operations to reduce processing time, and the readout and processing of a digital image sensor is

no exception; increasing the number of parallel output and oversample channels will decrease the readout time which currently dominates system timing and limits the sensor field rate and overall system frame rate. Chapter 7 describes an extension to this thesis work, which explores 'rolling' and adaptive field image oversampling permitting frame rate equal to the field rate in collaboration with research group colleague Dr Istvan Gyongy.

Figure 5.26 describes the FPGA firmware that performs the parallel oversampling operation. Each of the sixteen oversampling blocks deserialises one sensor output and processes one pixel per clock cycle. The bit selection continually increments every clock cycle (counting from 0 to 19 then resetting to 0), this selects which de-serialiser bit to read into the accumulator and which RAM address to write. Each pixel bit value of the field image is added to the previous value of the pixel from the RAM memory value. The memory lookup of the previous pixel value is one clock cycle ahead of the bit selection pipelining the operation and mitigating any redundant clock cycles. After processing the final field image capture, the memory controller flags that the frame data is ready and begins sequentially streaming the contents of the sixteen RAMs into an intermediary FIFO for readout via USB.

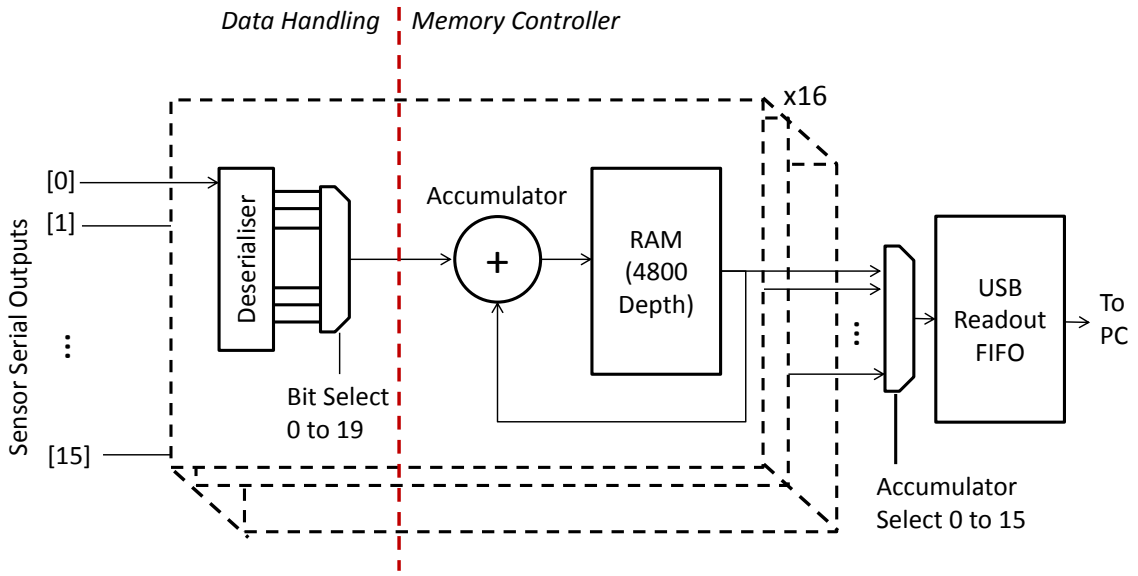


Figure 5.26. Block diagram of the FPGA firmware for digital sensor data handling. Sixteen parallel oversampling blocks consist of a de-serialiser, and a memory control block operating a sequential pipelined accumulation procedure.

5.4. Time of Flight

A pulsed ITOF depth image may be formed by capturing two, four, or more time-gated images (with different exposure times or delayed in time) synchronous to a nanosecond pulse laser and combining them through an image signal processing (ISP) algorithm. The sensor has two time-gate generation blocks which are employed here to capture two time-gated images in order to create a two-bin ITOF image.

The equation to calculate the distance to an object ('Z') using ITOF for a two-bin pulsed system from [66] is:

$$Z = c \cdot \frac{1}{2} \cdot TOF = c \cdot \frac{1}{2} \cdot \left[T_{PW} \cdot \left(\frac{A}{B} - 1 \right) - T_L \right]$$

(Eq. 5.5)

where T_{PW} is the pulse width of the laser illuminator, A and B are the two recorded intensities of the time-gated pixel values and T_L is the temporal offset of the laser. The B window captures the entire dynamic range of the TOF. To subtract the effect of ambient light, two images are captured without laser illumination. Equation 5.5 becomes

$$Z = c \cdot \frac{1}{2} \cdot TOF = c \cdot \frac{1}{2} \cdot \left[T_{PW} \cdot \left(\frac{A - A'}{B - B'} - 1 \right) - T_L \right]$$

(Eq. 5.6)

where A' and B' are the ambient subtraction values. To achieve this ITOF image capture with the image sensor, four sequential images must be captured. The time-alignment of the A and B windows, and the control of the four image captures, are described in the following sections.

5.4.1. Rising of Falling Edge Time Gates

In two bin ITOF image capture using photon counting, one bin (B) captures a gated window in time equal to the whole dynamic range [66]. The next bin (A) is temporally offset to capture only one half of the dynamic range. The distance is computed as the intensity ratio of the A bin value to the B value. In this situation, the time gate durations do not need to precisely equal each other or match the illuminator pulse width. Therefore the time gates are configured to be longer than the illuminator and the offset between them is set to be greater than or equal to the illuminator pulse width. There are two options for configuring time gates for ITOF image capture either using the rising or falling edge. The primary difference between the rising and falling edge time gate configurations is the detection of TOF reflections from targets in the 'dead zone' – the region in time (and therefore TOF distance) following the TOF range of interest. As shown in Figure 5.27 with a rising edge time gate the detected out of range time is:

$$T_{OR} = 2.\mathrm{T}_{PW} + (\mathrm{T}_{TG} - \mathrm{T}_{PW}) = \mathrm{T}_{TG} + \mathrm{T}_{PW}$$

(Eq. 5.7)

where T_{OR} is the detected out of range time, T_{TG} is the time gate width and T_{PW} is the illuminator pulse width. Whereas in a falling edge time gate configuration as shown in Figure 5.28, has the advantage in this regard as the detected out of range time and distance is reduced:

$$T_{OR} = \mathrm{T}_{PW}$$

(Eq. 5.8)

The time-gate falling edge configuration also benefits from not having the rising edge 'blur' described in Figure 4.2 and 4.3 where the recharge decay of the SPAD pulse is captured, which occurred some ns before the assertion of the time gate. In this falling edge configuration the ITOF equation is re-arranged to:

$$Z = c \;.\; \frac{1}{2} \;.\; \left[T_{PW} \;.\; \left(1 - \frac{A - A'}{B - B'} \right) - T_L \right]$$

(Eq. 5.9)

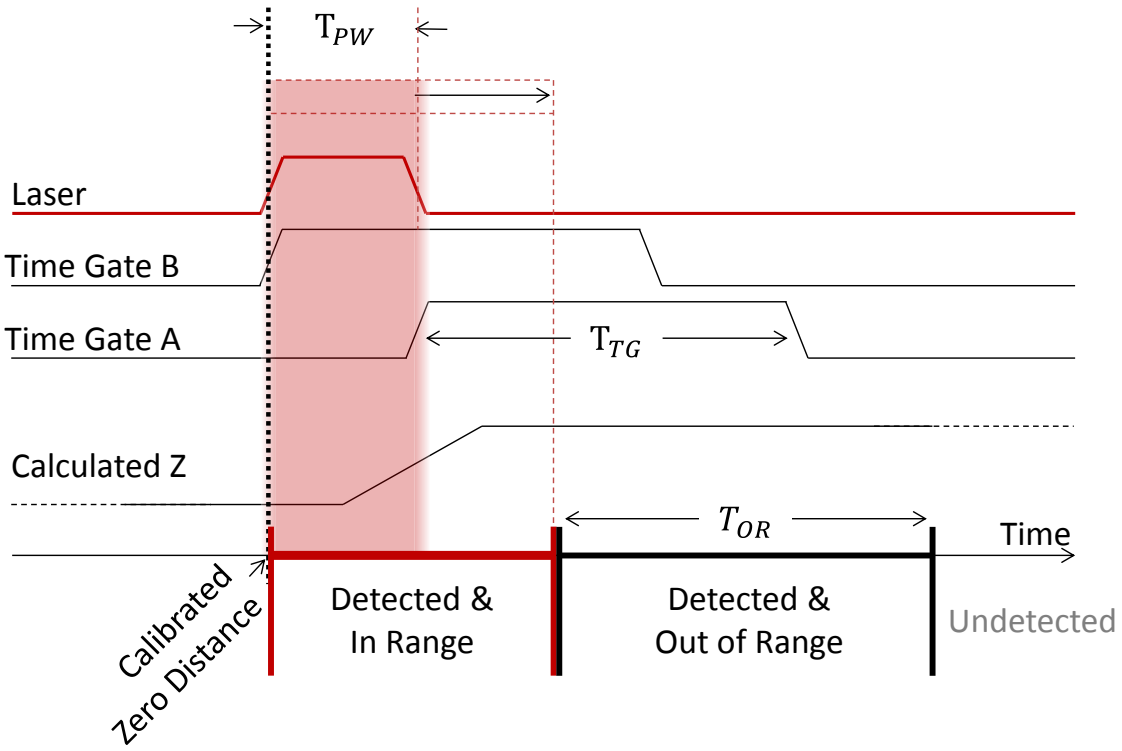


Figure 5.27. Pulsed ITOF time-gate rising edge configuration.

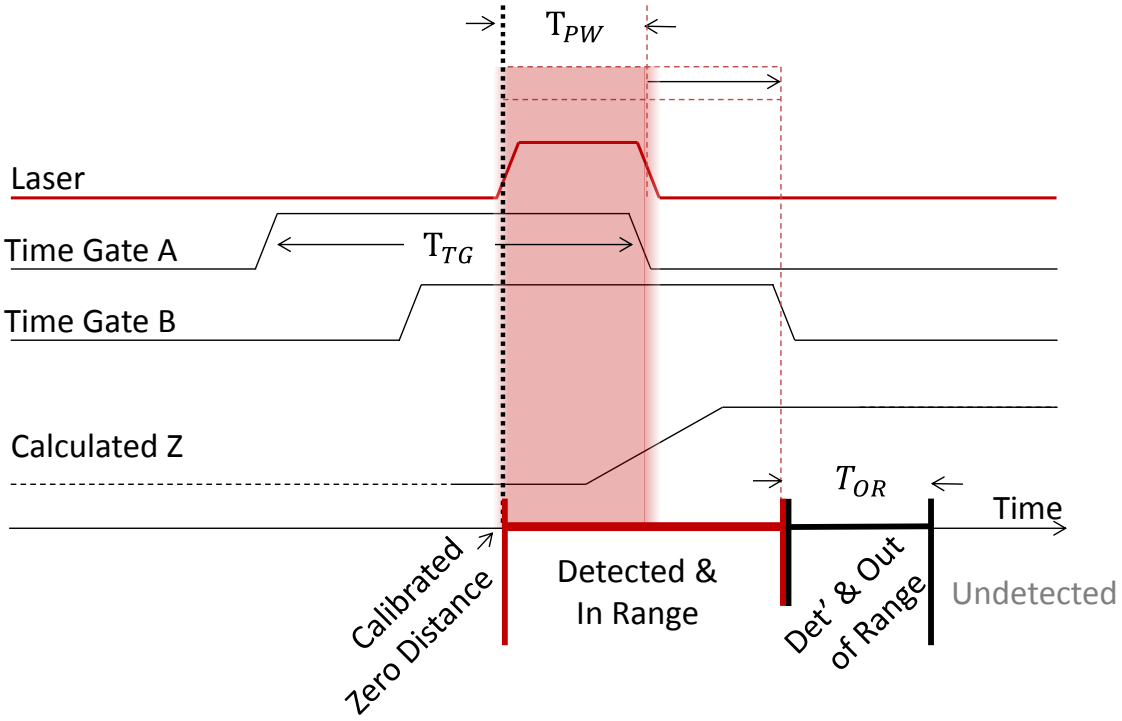


Figure 5.28. Pulsed ITOF time gate falling edge configuration has a reduced out of range detection to the rising edge configuration.

5.4.2. TOF Sensor Timing

The number of photons captured in a single nanosecond time-gated photon counting image co-incident with an illuminating laser flash is assumed to be low. To increase the signal received on the sensor before readout, the illuminating pulse and time gate triggering is repeated. As a number of time-gated images are integrated in the pixel, the event under observation must be repeatable and the imaged subject must have no relative motion between integrations. To perform this integration on the sensor, the illuminator pulse, the time gate input and the time gate disable are toggled. All external laser drivers have some temporal offset 'T_L' (from receiving electrical stimulus to optical transmission) which must be manually matched or initially calibrated by the use of the on-chip delay lines. This toggling and temporal offset of time-gate, time-gate disable and illuminator pulse are illustrated in Figure 5.29 for falling edge time-gate configuration.

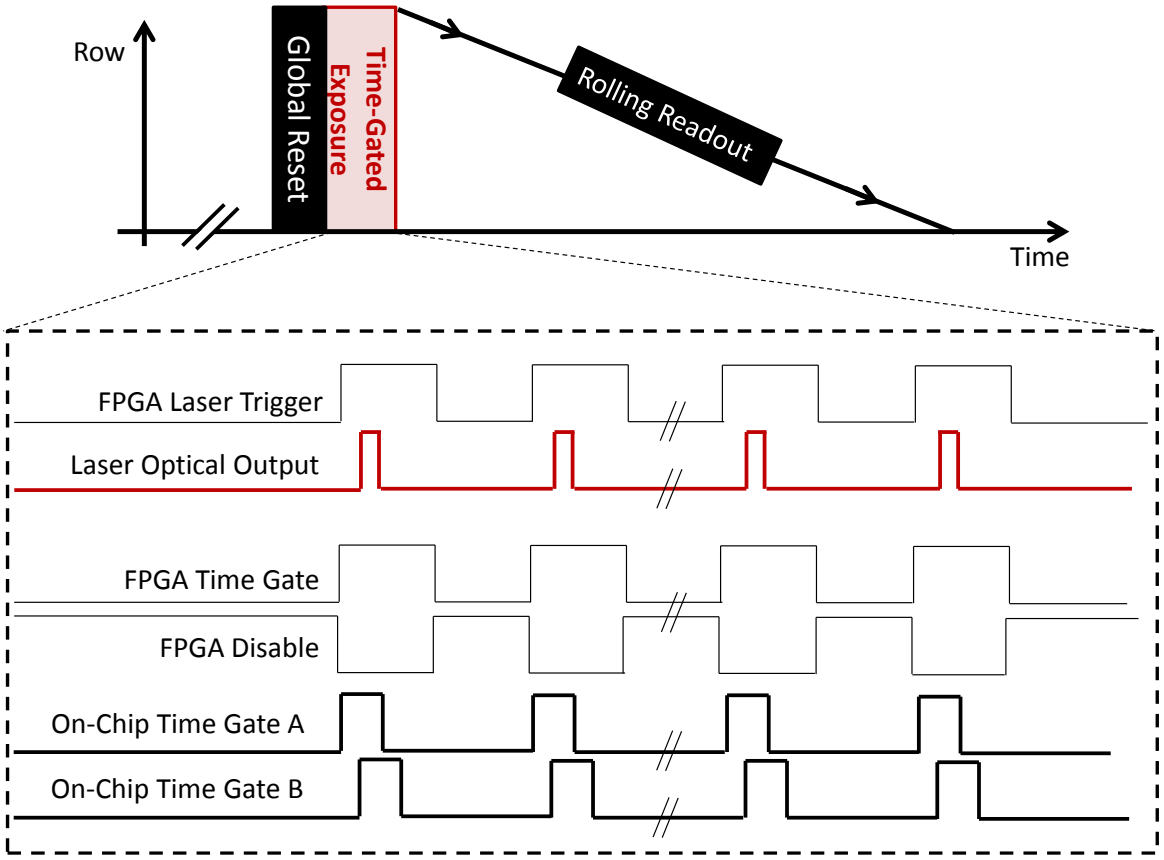


Figure 5.29. Sensor timing diagram for time-gated image capture synchronised with a repeating laser optical stimulus.

5.4.3. Sequential or Interleaved Time Gates

The pulse generators (for A and B time-gates) are configured separately to have two differing time gates, allowing the sensor to capture two time-delayed images in succession. Each row time gate selection mux is reset to select pulse generator A, and every subsequent activation of the row read signal toggling the time gate input. As such, by selecting particular rows before an exposure, a pattern may be configured whereby the sensor captures two interleaved time-gated images simultaneously. Three envisaged time gate selections are shown in Figure 5.30, the first image (a) shows standard sequential operation, and the second and third images (b) and (c) show interleaving options if 2x2 binning is in place. The disadvantage of sequential capture in (a) is the movement of the imaged scene during successive iterations causing distortions in the computed TOF. Double interleaved rows shown in Figure 5.30 (b) seeks to address that by capturing the two time gated images in parallel. The additional motivation for (c) is to half the required frame rate with binning in place.

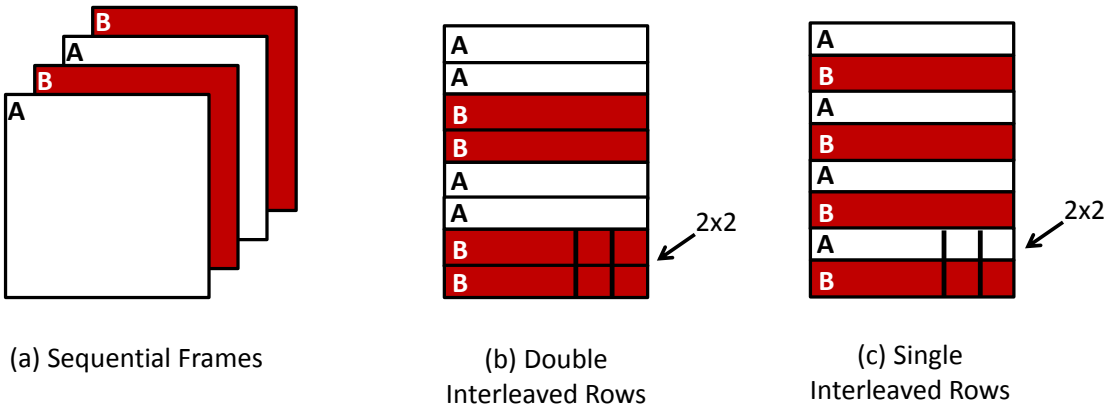


Figure 5.30. Sequential or interleaved time gating.

5.4.4. Analogue Counting ITOF Image Capture

Four field images are sequentially captured to perform ITOF with background subtraction. The first two capture the ITOF images for bins A and B (TOF A and TOF B). In the next phase, ambient subtraction is performed by disabling the illuminator and capturing the background (BG) field images (BG A' and BG B'). The exposure of each image is adjustable which also changes the number of laser repetitions per TOF image. The background images are subtracted from the TOF images to create a depth map in post-processing software.

5.4.5. Digitally Oversampled Time of Flight

This section describes utilising the time-gated binary image sensor mode to create oversampled ITOF images. The parallelised FPGA memory and accumulator blocks are dynamically re-configured to build two images (Bin A and Bin B) side by side in real-time. Figure 5.31 displays the timing diagram of one digital ITOF exposure capturing these four bin images. The two field images in the 'TOF + Count Up' phase are added to the previous field images oversampled in memory. Then the two field images in the 'BG – Count Down' phase are subtracted from the memory value. This process is repeated, summing up a number of TOF fields and subtracting ambient fields, to build up the digital TOF frame. By capturing the four fields together in an interleaved fashion, instead of capturing four frames one after the other, this ensures that motion blur equally affects all bins and does not result in a momentary TOF frame range error.

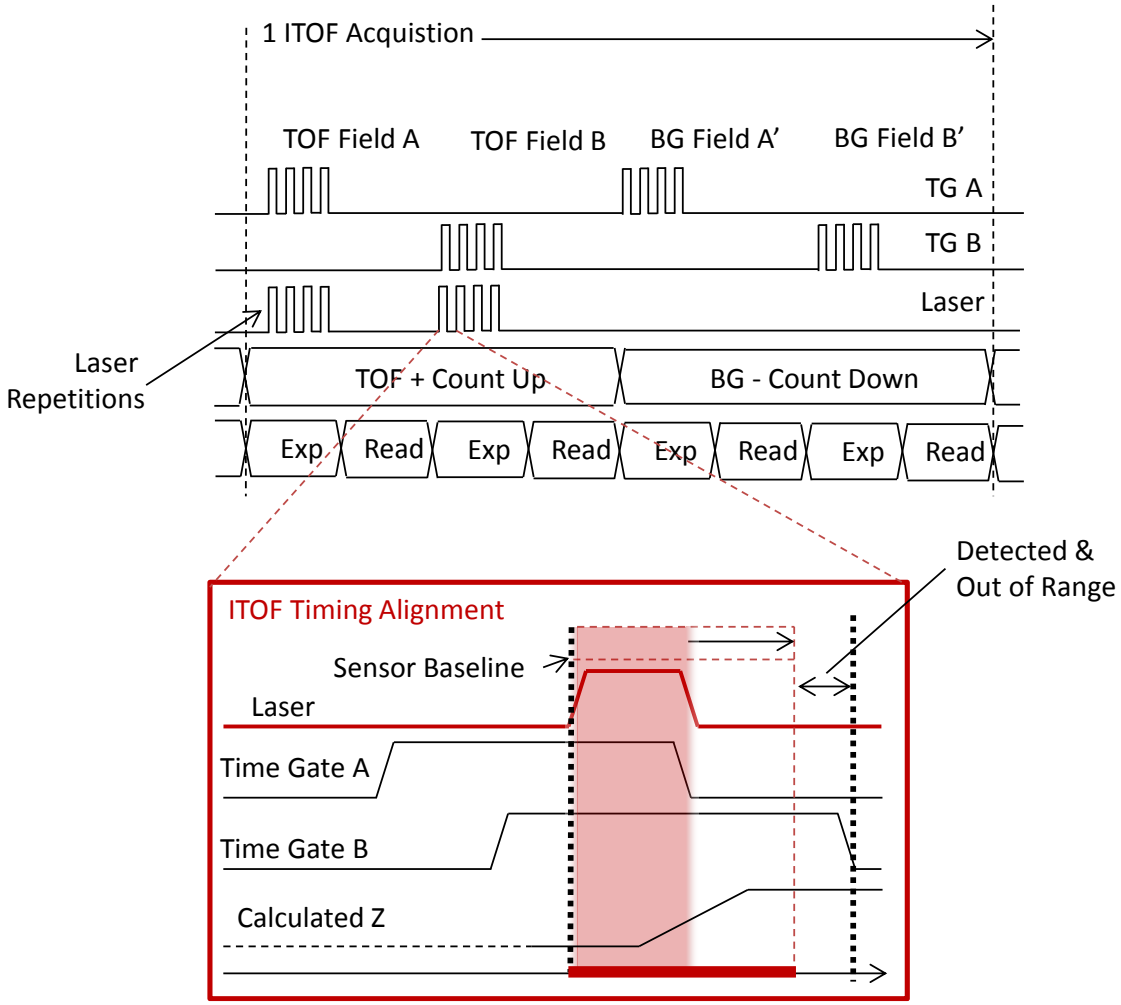


Figure 5.31. Digital ITOF timing diagram of four field images.

Like the analogue TOF capture, each field image consists of a user defined exposure time and consequently number of illuminator repetitions. However, the process is limited by the maximum accumulation memory width available – the digital or QIS equivalent of full well. On the particular Xilinx Spartan 3 FPGA in use,

the maximum permissible is 18b or 9b per bin image and a maximum of 512 TOF fields to avoid clipping. The readout from the FPGA is fixed at 16b, so the LSB of each bin 9b TOF frame image is dropped and the output bit-shifted in software. This quantisation will increase the distance noise in the final TOF image.

5.4.6. TOF ISP Algorithm

A basic algorithm is developed to compute a depth map from the four captured images and implemented in MATLAB software. There are two parts to the algorithm: thresholding and division. The basis of the algorithm is shown as a flow chart in Figure 5.32 assuming bin B captures the image of the whole TOF dynamic range and bin A is temporally offset (as previously described). The threshold algorithm suppresses low intensity regions of the images which will have high range noise.

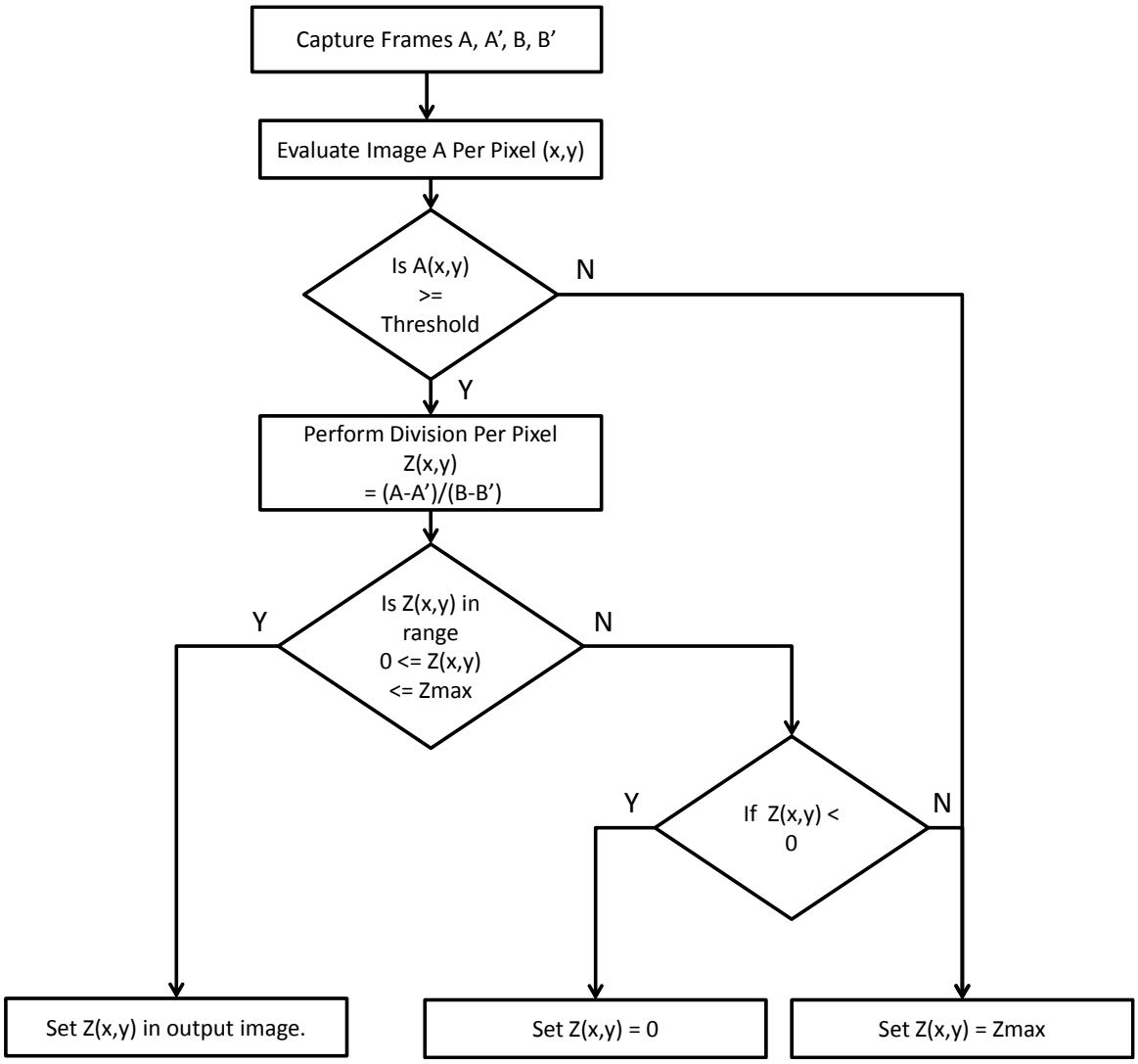


Figure 5.32. Basic TOF algorithm flow chart.

5.5. Summary and Discussion

This chapter has described the implementation of a 320x240 SPAD-based single photon counting time-gated image sensor. The initial specifications proposed for the imager were discussed.

The proposal for SPAD-pixel design discussed in Chapter 3 was implemented in the form of an 8um pitch, 26.8% fill factor, dual in-line strip format SPAD array. The original test pixel electronics from the 3x3 array in Chapter 4 were revised, removing two transistors, forming a 9T NMOS-only pixel. At the time of publication of [2], the revised pixel design was the state of the art for a SPAD-based image sensor in terms of number of transistors, pitch and fill-factor. The physical pixel layout and array metallisation were described.

Two revisions (original 'AA' and a metal fix 'AB') of the IC were fabricated. The former has time-gate drivers from both sides of the array and the latter only from one side as no compensation was included on-chip to counterbalance for temporal mismatch across the two halves of the array.

The row and column logic for sensor operation were described and the control functions and signal timings were illustrated. The sensor has two readout chains serving two distinct modes of image capture. Utilising these, four modes of image capture are proposed. Single photon intensity images or time-gated images are captured either by analogue counting or digital oversampling. The analogue readout operating at approximately 2.5FPS, permits single photon sensitive global shutter images or analogue counting ITOF images to be captured. On the other hand, the single bit digital readout combined with FPGA-based oversampling operates at >5kFPS. This binary image capture will permit exploration of the Quanta Image Sensor concept.

The analogue read-out timing describes the use of CDS delta reset sampling and read-out crowbar DDS sampling. The use of delta reset over true CDS sampling will increase the image noise but with the advantage of capturing global shutter images.

Time gating options with rising and falling edge timing configurations were discussed. Falling-edge is preferred for two reasons: the rising edge suffers from temporal 'blur' capturing SPAD events occurring before the onset of the time gate, and the reduction in capturing 'out of range' TOF targets. The novel concept of using the QIS mode to capture interleaved gated binary images to form digitally oversampled TOF depth maps is described.

For the capture of an ITOF image, the two different pixel operation and readout techniques have certain advantages and disadvantages when scaled to a higher resolution array. Both techniques require oversampling into a frame store to achieve a similar full well to commercial ITOF sensors based on photo-gate or photo-diode. Table 5.6 seeks to compare the two modes in terms of an ITOF capture requiring four ITOF frames With the integration in-pixel the analogue technique requires 80x less frame reads (for an example full well of 80 SPAD counts) than the digital technique. However as the digital mode is implemented with column

parallel readout (instead of the analogue sequential readout) it recovers this by high output frame rate for similar levels of power consumption.

	Analogue	Digital
Full Well (Max Count)	80	1
Time Gated Image Frame Rate (FPS)	3.5	5000
Oversampled Frames per ITOF Bin Image	10	800
Total Number of Frames Required	40	3200
ITOF 4 Bin Frame Rate (FPS)	0.09	0.64

Table 5.6. ITOF capture mode comparison for the QVGA image sensor described in this research.

Table 5.7 presents an alternative scenario if the image sensor was scaled four times larger to VGA (640x480), and the analogue mode was designed with column parallel readout (8b ADC) with four times the data readout rate of the QVGA sensor. Again, the power of each mode would be similar. The digital mode frame rate does not scale to higher resolution arrays when implementing off-chip readout and is a key reason to move to chip stacking technology for on-chip oversampling.

	Analogue	Digital
Full Well (Max Count)	80	1
Time Gated Image Frame Rate (FPS)	160	5000
Oversampled Frames per ITOF Bin Image	10	800
Total Number of Frames Required	40	3200
ITOF 4 Bin Frame Rate (FPS)	4	0.64

Table 5.7. ITOF capture mode comparison for a scaled up VGA image sensor with on-chip column parallel ADCs.

6. SPAD and Image Sensor Evaluation

This chapter provides characterisation data, measurement results, analysis and evaluation of the image sensor and the accompanying SPAD test structure IC.

The SPAD is measured using a test chip for characterisation of breakdown voltage, dead time, maximum count rate, temporal resolution or jitter, after-pulsing, current consumption, and using the image sensor for DCR.

The image sensor is demonstrated with a variety of imaging modalities: single photon counting imaging using analogue counting, oversampled binary or QIS image capture, time-gating performance and ITOF imaging in the QIS digital mode.

6.1. SPAD Performance

A small 12 pad test-chip (*'UNIED_SPAD_WLC_BA PND AA/AB'*) was fabricated on the same MPW as the image sensor (*'UNIED_SPCIMAGER_AA/AB'*). The test chip parts measured are from the same wafer as the image sensor revisions 'AA' and 'AB'. The test IC contains eight discrete SPADs from the image sensor that are each individually routed out through an output buffer (six stage inverter chain) to a pad. The SPADs in this test-chip are used to evaluate the SPAD performance in terms of breakdown voltage, dead time, count rate, SPAD current consumption, after-pulsing, photon detection probability (PDP) and temporal jitter. The SPAD array of the image sensor is used to characterise the SPADs in terms of DCR. Figure 6.1.shows a photomicrograph of the 12-pad test-chip with a zoom inset of the 4 by 2 SPAD array. The wire-bonded output pads are visible on the left hand side of the main image.

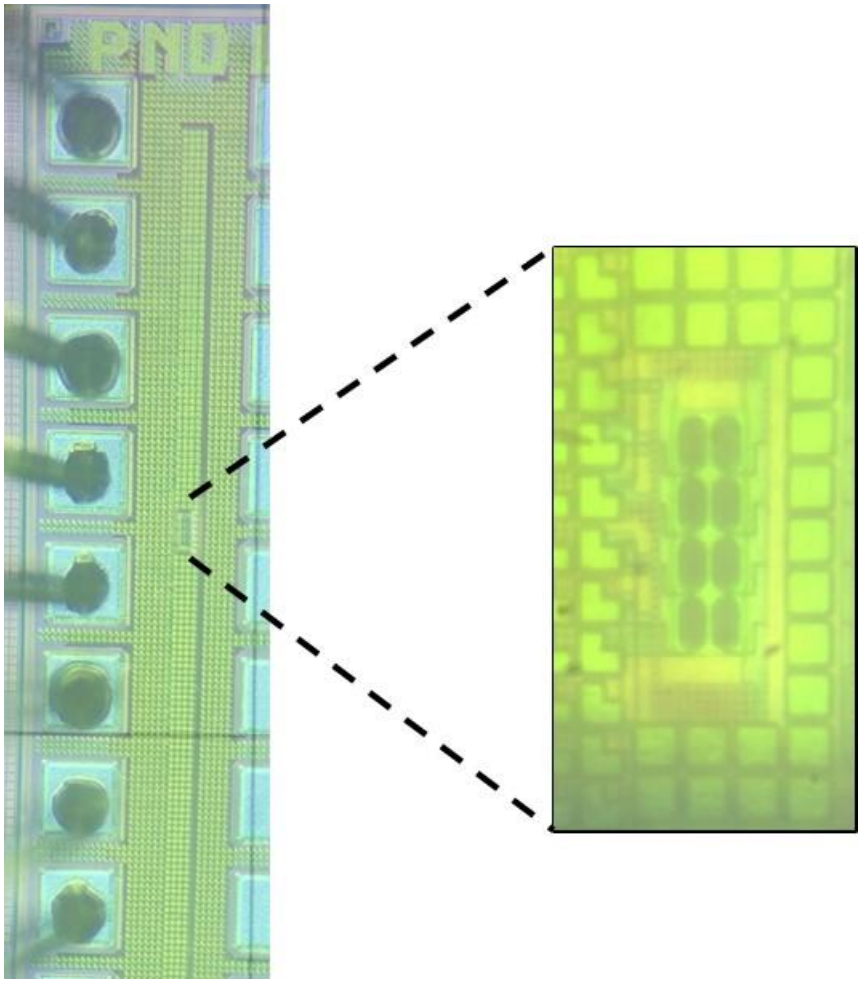


Figure 6.1. Photomicrographs of the 'UNIED_SPAD_WLC PND AB' 12-pad test chip used for discrete SPAD evaluation.

6.1.1. Breakdown Voltage

The breakdown voltage was measured across five test IC 'AB' samples under bright constant light from a commercial 485nm LED powered by a DC supply, monitoring the output of the SPAD buffers on a LeCroy oscilloscope. The mean onset of recording SPAD pulses is 13.80V. Assuming matched PMOS and NMOS devices in the front end inverter, the threshold voltage of the inverter is 0.6V with a 1.2V power supply. This indicates the mean breakdown voltage is 13.20V at room temperature (22°C). Table 6.1 states this and the excess bias equivalent voltages used throughout this chapter.

Mean Breakdown Voltage at 22°C	13.20V
1V Excess Bias Equivalent	14.20V
2V Excess Bias Equivalent	15.20V
3V Excess Bias Equivalent	16.20V

Table 6.1. Mean breakdown voltage and equivalent excess bias voltages.

6.1.2. Dead Time Modelling

To model the dead time of the SPAD against quench voltage, the DC resistance of the NMOS PQPR is simulated with an equivalent circuit of the SPAD. The simulated equivalent circuit is shown in Figure 6.2. The dead time of the equivalent circuit is calculated, using a transient simulation, as the time of the anode voltage above a 0.6V threshold (equivalent to an inverter threshold at a 1.2V supply). The excess bias of the SPAD equivalent circuit was varied for three voltages (1V, 2V and 3V). The quench voltage of the PQPR transistor is incremented from 0.1V to 3.3V in 0.1V steps. The current through the piece-wise current source was optimised to create the desired excess bias voltage swing on the anode using a 1ps switch 'closed' period to represent a SPAD avalanche. The capacitance value is modelled on the value of the diode P-well capacitance.

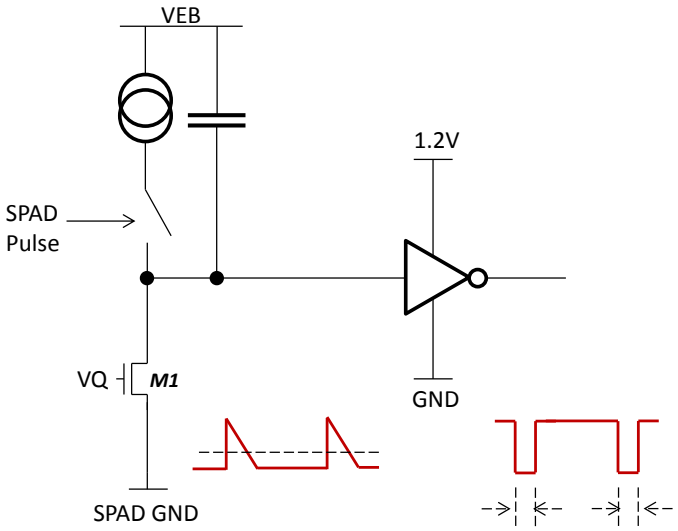


Figure 6.2. Simulated SPAD equivalent circuit to evaluate quench resistance versus dead time.

The dead-time simulation results are shown in Figure 6.3. The resistance of the PQPR transistor (0.97µm length, minimum 0.3µm width NMOS) is also plotted in Figure 6.3 illustrating the inverse relationship of gate voltage to resistance. Bumps or lack of smoothness in the simulated curves are due to simulation quantisation or calculation errors. The rule of thumb of 50kΩ per 1V of excess bias from [93] is plotted alongside with three coloured dotted lines indicating the quench voltage required at each of the excess bias settings.

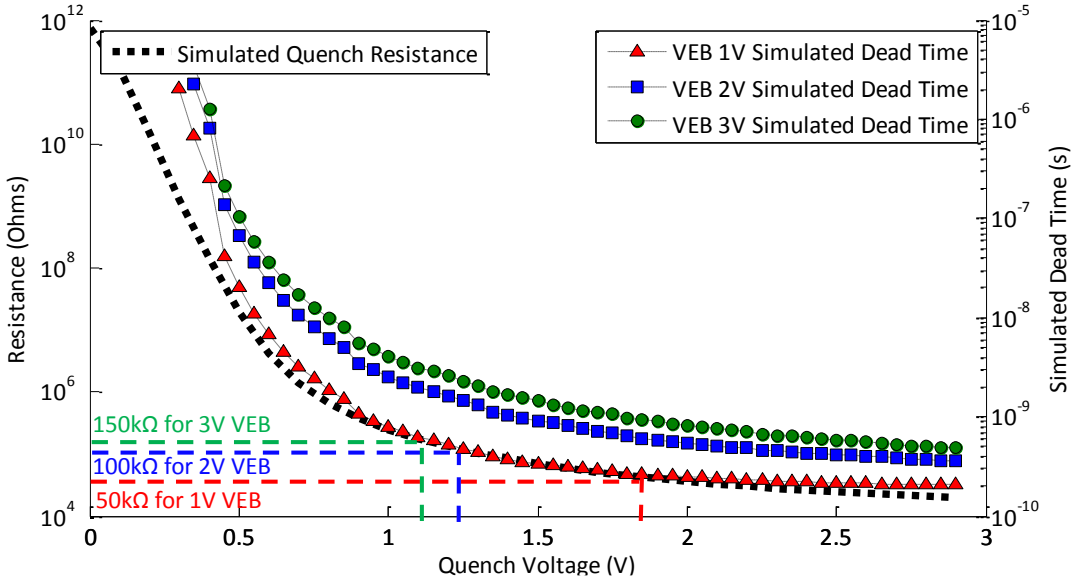


Figure 6.3. Graph of simulated single transistor NMOS quench resistance versus quench voltage, with the simulated SPAD dead time of three excess bias conditions extracted at a fixed threshold of 0.6V.

6.1.3. Dead Time Measurement

The dead time of the SPAD was characterised using the pulse width measurement function on a LeCroy 10GS/s 'WavePro' oscilloscope with 1.5GHz bandwidth probe. The 'AB' test-chip was illuminated with constant illumination from a commercial 485nm LED with regulated DC supply. The experiment was conducted measuring SPAD dead time versus quench voltage and repeated for three excess biases (1V, 2V, 3V). The upper graph in Figure 6.4 presents the plotted results of mean dead time versus quench voltage (VQ) for the three excess biases. The lower graph in Figure 6.4 displays one standard deviation of the same data against VQ. According to the basic quench resistance model, the dead time should plateau for 1V< VQ ≤3.3V. However, it is clear that in the region VQ>1.5V, the mean and standard deviation of the dead time significantly increases and eventually the output paralyses and 'locks up' high. Table 6.2 shows the maximum quench voltage at each excess bias condition at which the dead time variability is minimised. This is critical as variations in dead time will induce counter variability and PRNU in an image.

Figure 6.5 shows the standard deviation normalised against the mean, which reveals that there are four populations of behaviour. Weak inversion operation matches simulation with long dead time for VQ in the region of V_{TM1} which operates the M1 PQPR NMOS in weak or moderate inversion. Threshold voltage variation in weak inversion creates large variations in device resistance and the >10% variation in dead time is expected. Reliable SPAD operation is encountered with 1.0V≤VQ≤1.5V (for all three measured excess biases) with M1 in strong inversion, which is line with the dead time model. The last two regions of SPAD behaviour present unpredictable behaviour in dead time as the standard deviation becomes greater than the mean. A

hypothesis is that as the SPADs count rate increases, the cathode NW is highly resistive and cannot recharge the deep NW. It is clear that the rule of thumb for quench resistance proposed in [93] is experimentally verified here for this SPAD. This phenomenon warrants further investigation.

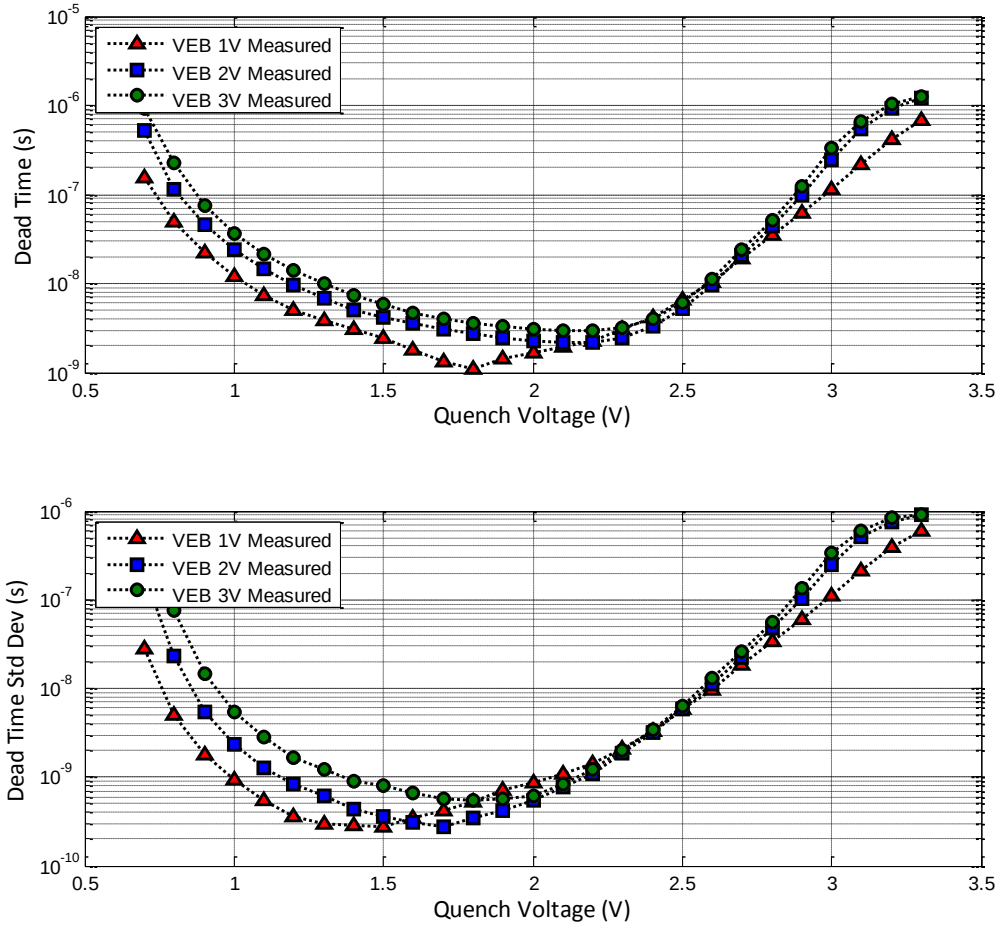


Figure 6.4. Measured test IC SPAD dead time, plotted as the mean and one standard deviation versus quench voltage at room temperature.

Excess Bias	Maximum Quench Voltage Before Unpredictable Operation
1V	1.5V
2V	1.7V
3V	1.9V

Table 6.2. The optimum quench voltage at each excess bias condition to minimise dead time variability.

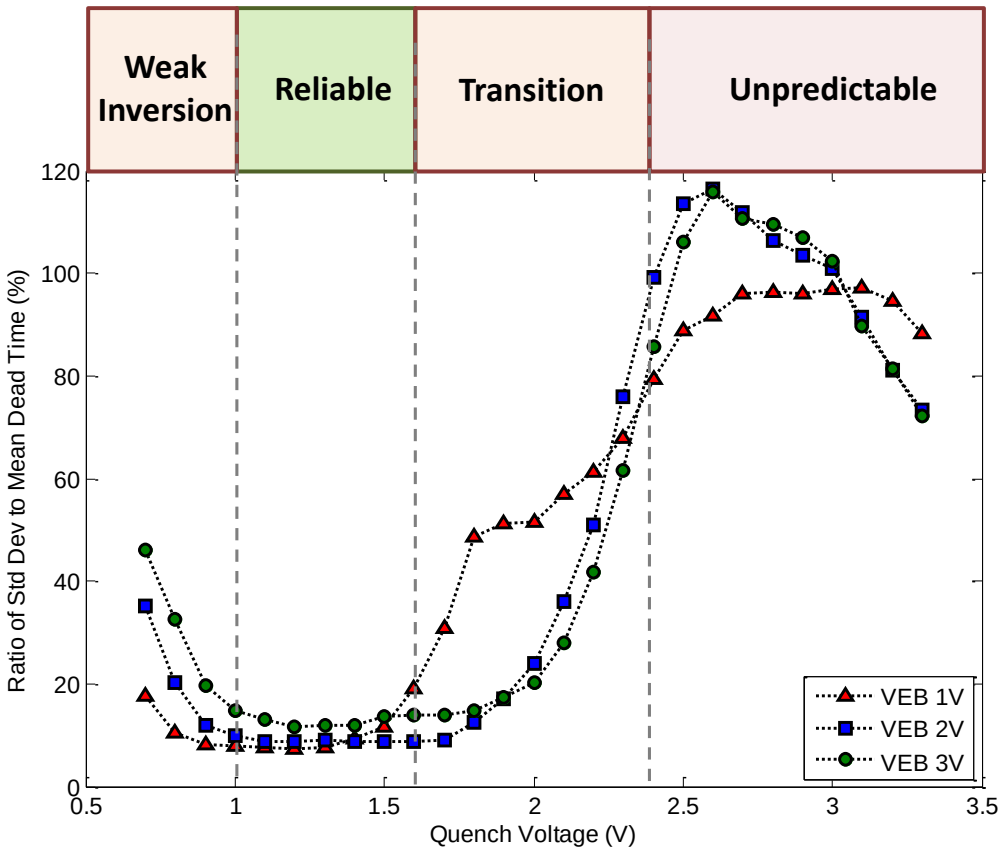


Figure 6.5. Plot of the ratio of standard deviation versus mean dead time with four regions of operation highlighted.

6.1.4. SPAD Count Rate and Current Consumption

The count rate and current consumption of a single SPAD from the 'AA' revision test IC array was measured versus illumination level. The count rate (in counts per second (CPS)) was measured using a LeCroy 'WavePro' oscilloscope with 0.9pF 1.5GHz active probe, the current consumption was recorded using a Keithley power supply, and the illumination level was controlled by a customised calibrated 'bright light' experimental setup (devised by the ST characterisation team). The latter encompasses a temperature-controlled bright light source directed through a controllable mechanical iris into an illumination sphere with discrete Newport 1830C/818-UV photo-diode (in one of the two sphere ports) for optical feedback, with the system automated by MATLAB. The SPAD test IC was then aligned at the second sphere port.

The recorded count rate data are plotted against increasing illumination power in Figure 6.6, and the current consumption, at 1V EB, in Figure 6.7. The flat region of count rate and current, with no relationship to the light level, is due to DCR. Both SPAD and image sensor were 'AA' revisions, before DCR reduction by process improvement was applied in the 'AB' revision. The 'bright light' setup could reach a maximum of 1k Lux with the sensor positioned 7cm from the sphere port. The onset of SPAD paralysis is not observed. The measurement was repeated using the image sensor and the recorded current is shown in Figure 6.8. The image

sensor count rate could not be measured, due to a conflict of communications protocols in the experimental setup. Extrapolating from the current and count rates graphs, a comparison is made to evaluate the current or charge per SPAD event. The extrapolated count rates, currents and charges are described in Table 6.3. The mean charge per SPAD event in the test IC is 0.03pC per SPAD pulse, and the extrapolated mean in the image sensor is 0.012pC per pulse. A hypothesis is made that the image sensor current per SPAD pulse is lower than the small test structure suggesting that the image sensor VHV supply has higher resistivity (whether through supply lines or through cathode NW diffusion resistance) and limits the count rate of the image sensor. A firm conclusion is drawn from these data, for both discrete SPAD structure and image sensor SPAD array, the current consumption per SPAD event falls for increasing light level. This suggests that subsequent SPAD avalanches occur before being the diode is fully recharged at high light levels.

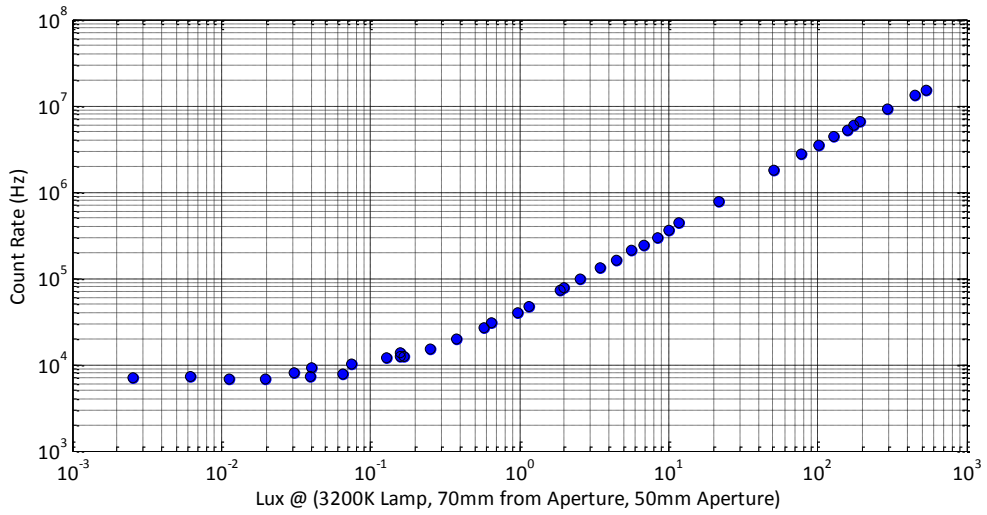


Figure 6.6. Single SPAD count rate against illumination power at room temperature.

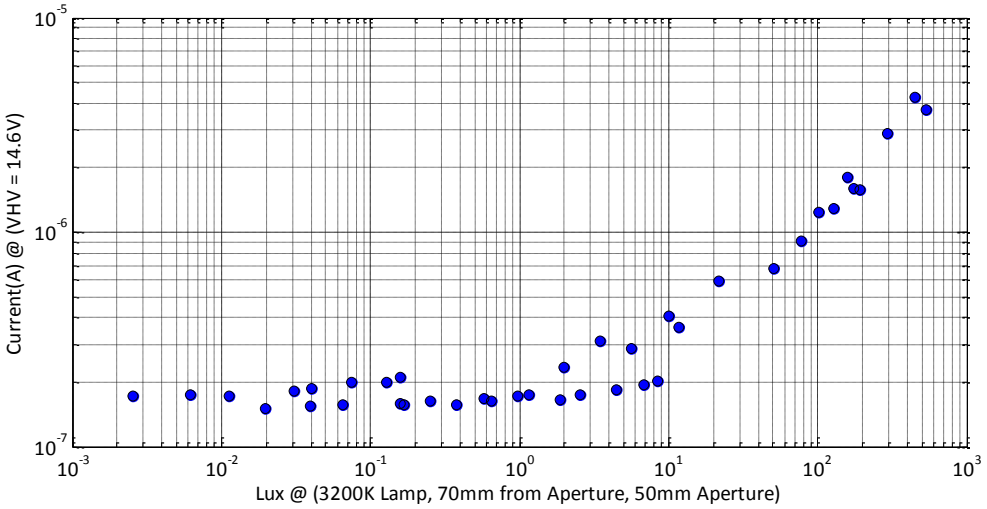


Figure 6.7. Single SPAD current consumption against illumination power at room temperature.

It is apparent from the 10kLux scenario that the 100mA current consumption is untenable for mobile applications. Possible solutions to reduce this current are optical notch filters to significantly reduce ambient outside the bandwidth of interest, reduced excess bias (at the loss of sensitivity), and fast system response in activating the SPAD array only for the required integration time. Thermal management also becomes a priority at these high current levels.

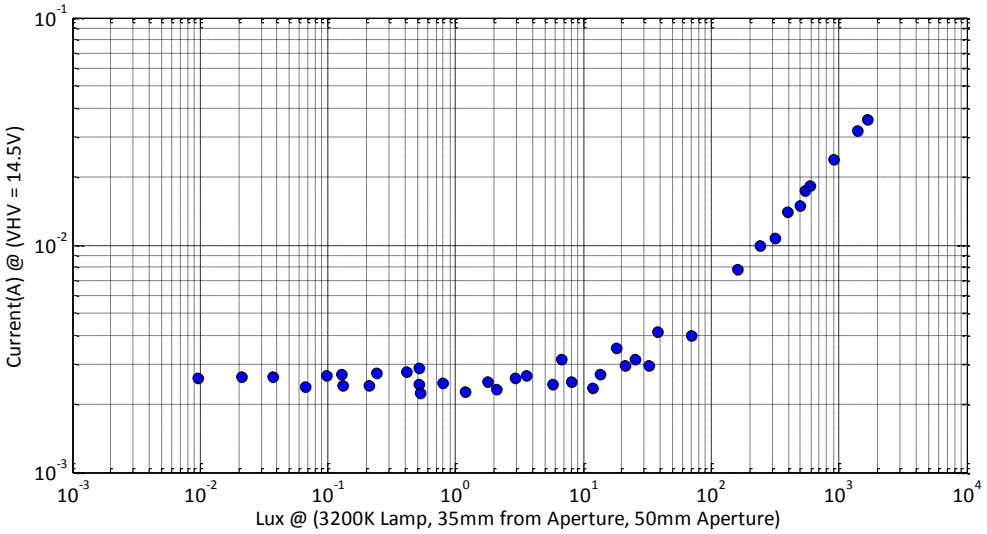


Figure 6.8. 'SPCIMAGER_AA' 320x256 SPAD array current consumption against illumination power at room temperature.

Lux	100	1,000	10,000
Test IC Extrapolated Count Rate (CPS)	3.1×10^{6}	30×10^{6}	290×10^{6}
Test IC Extrapolated Current (A)	1.2×10^{-6}	6.8×10^{-6}	43×10^{-6}
Test IC Current Per SPAD Event (pA / CPS)	0.0048	0.0028	0.0019
Image Sensor Extrapolated Current (A)	5.5×10^{-3}	25×10^{-3}	102×10^{-3}
Estimated Charge Per SPAD Event (pC / SPAD event)	0.0022	0.0010	0.0004

Table 6.3. Extrapolated SPAD charge, SPAD counts and SPAD VHV current consumption

6.1.5. Temporal Jitter

The temporal jitter of the SPAD was measured using a PicoQuant PDL-800 laser driver and 425nm laser head with 80ps-FWHM reported pulse width, and LeCroy 'WavePro' oscilloscope with 1.5GHz active probe. The electrical synchronisation (sync) from the driver triggered the scope, and the edge-to-edge jitter (sync to SPAD pulse) was integrated and recorded for 1,500,000 counts. Figure 6.9 shows both the linear and logarithmic histograms of the integrated SPAD jitter. No significant jitter tail is recorded. The standard deviation of the integrated jitter is calculated at 78.4ps, and the FWHM is calculated at 184.6ps. Subtracting the quoted laser FWHM jitter value in quadrature, the FWHM of the SPAD jitter is calculated at 166.3ps.

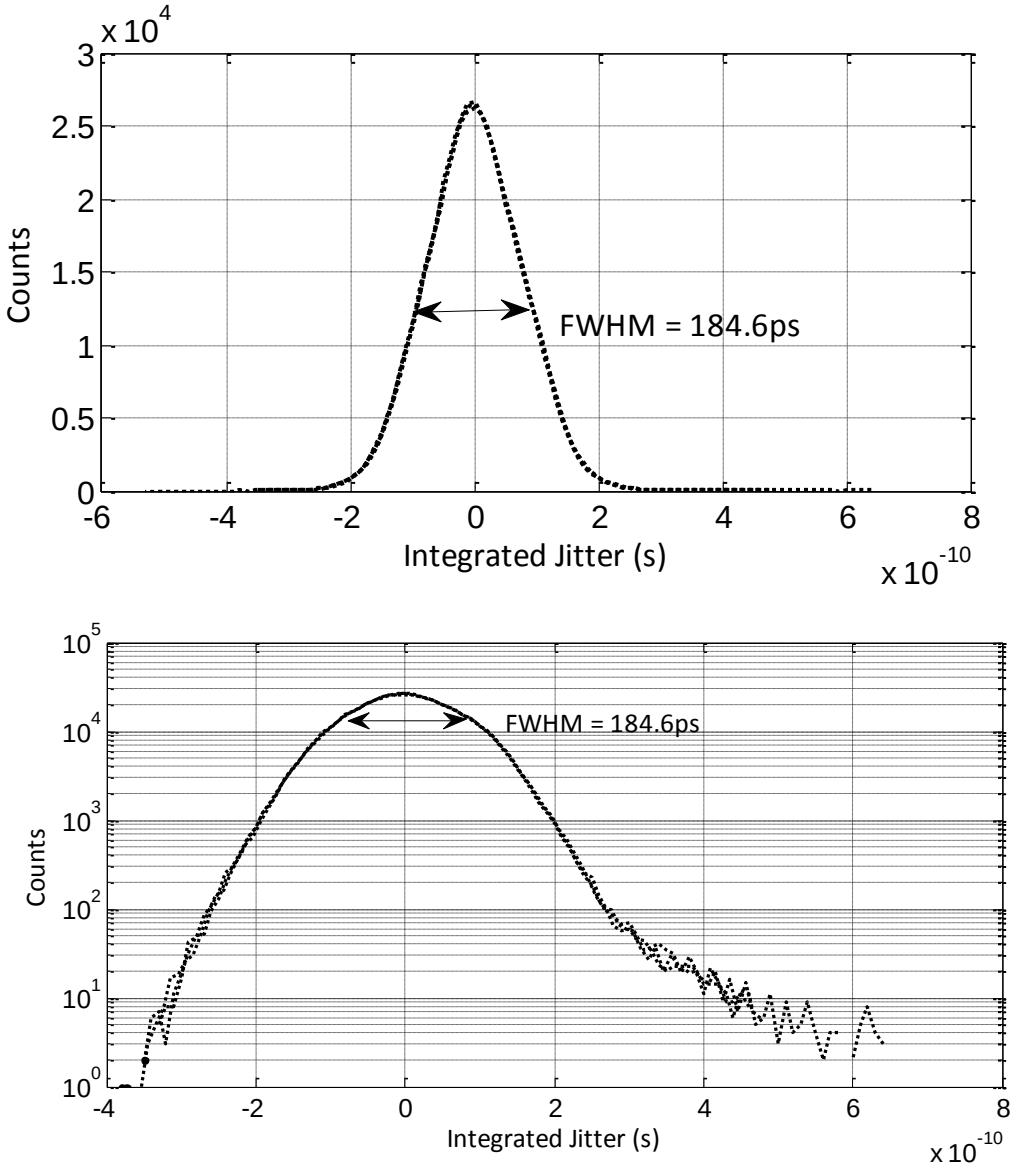


Figure 6.9. Linear and logarithmic histograms of measured integrated jitter of SPAD test structure optically triggered using a PicoQuant FDL-800 driver with 80ps-FWHM 425nm laser head.

6.1.6. Photon Detection Probability

The photon detection probability (PDP) is the SPAD equivalent of intrinsic QE (IQE) and the photon detection efficiency (PDE) is the equivalent of extrinsic QE which takes fill factor into account [68]. The results presented here, are measurements of PDP only. To measure the PDP, the number of SPAD output pulses are recorded using an oscilloscope at specific wavelengths across a spectral range. The PDP is calculated as the ratio of recorded counts to the theoretical number of incident photons on the diode active region.

The sweep is performed using a calibrated experimental setup comprising mono-chromator (Sciencetech 9055) with reference feedback photodiode (Newport 1830C/818-UV) in dark room conditions. The digital outputs of four SPADs in the test structure IC, are probed using a LeCroy 'WavePro' oscilloscope in a 40ms integration window with 1.5GHz active probes. The experiment is repeated for three excess biases (1, 2, and 3V) across the range 350nm to 1050nm in 10nm increments. At each repetition, the count rate is measured in the dark (DC) and under the mono-chromator light source (light counts (LC)) to track for fluctuations in DCR due to temperature or voltage. The number of incident photons is calculated per wavelength using the following Equation 6.1:

$$N(\lambda) = P_{Ref\_PD} \,.\, Norm_{Ref\_PD} \,.\, A \,.\, \frac{\lambda}{h \,.\, c}$$

(Eq. 6.1)

where N(λ) = number of photons incident on the SPAD active region, P_{Ref_PD} is the measured optical power on the photodiode, $Norm_{Ref\_PD}$ is the photo-diode normalisation factor (correction for photo-diode area and wavelength), A = SPAD active region area (calculated from PW area), λ = wavelength, h = Planck's constant, c = speed of light in a vacuum. The PDP per wavelength is calculated in Equation 6.2 as a percentage of incident photons:

$$PDP(V_{EB}, \lambda) = \frac{LC(V_{EB}, \lambda) - DC(V_{EB}, \lambda)}{t_{int}} \cdot \frac{1}{N(\lambda)}$$

(Eq. 6.2)

where both LC and DC are measured for wavelength λ and excess bias V_{EB}, in integration time t_{int}. The calculated PDP is plotted versus wavelength in Figure 6.10. The peak PDP is at 480nm (blue to green) and 5% PDE at 850nm NIR. This is in line with previous results from [92]. Table 6.4 lists the peak PDP and three NIR wavelengths of interest. For reference, the equivalent PDE of each of the listed figures is 3.7 times lower assuming the area of the P-well is equal to the photosensitive region (i.e. not including any perimeters effects), to a first order this is considered an acceptable approximation as the high electric field in a typical device simulation is shown to end within less than 100nm of the P-well boundary.

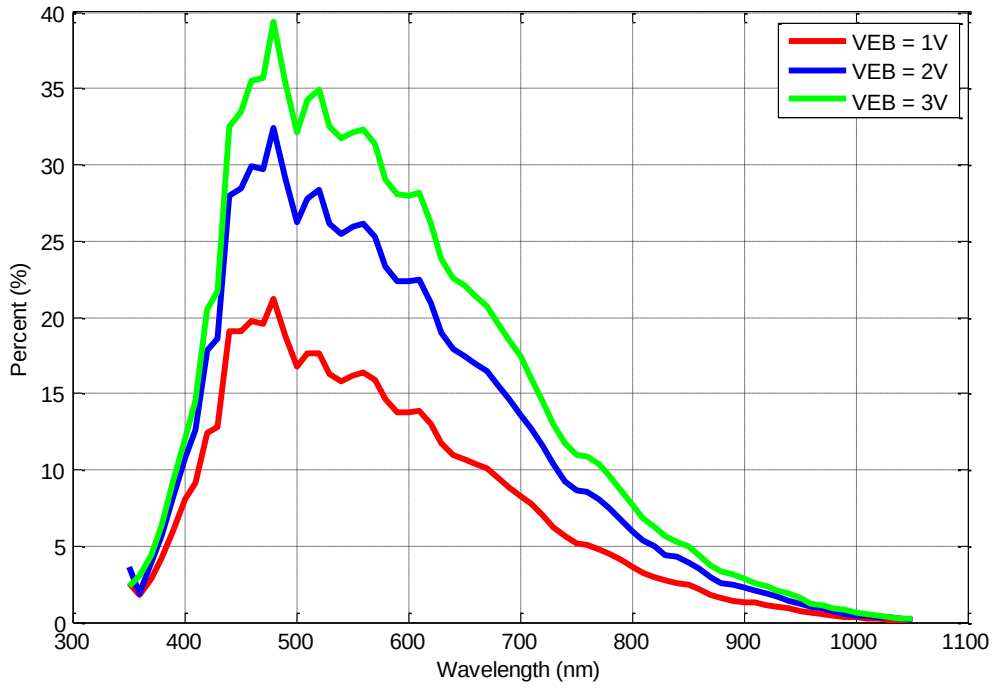


Figure 6.10. PDP graph of the UNIED_SPAD_WLC_BA PND 'AB' SPAD. The PDP was measured at three excess bias conditions (1V, 2V and 3V).

	Excess Bias		
	1V	**2V**	**3V**
Peak PDP @ 480nm	21.2%	32.4%	39.4%
PDP @ 840nm	2.6%	4.3%	5.3%
PDP @ 850nm	2.4%	4.0%	5.0%
PDP @ 940nm	0.9%	1.4%	1.9%

Table 6.4. PDP reference table showing peak PDP and three NIR wavelengths of interest.

The PDP graph shows the photo-sensitivity does not increase linearly from 2V to 3V excess bias. This is expected and in line with prior results for this SPAD structure [163], indicating that the probability of avalanche is asymptotically approaching 100% above 3V excess bias.

6.1.7. DCR

The DCR of the 'AB' chip revision is measured by using the digital oversampling mode with a short 500ns binary field integration time, with 2^{16} field exposures summed in the FPGA per frame and then 10 frame exposures summed in post processing. This is equivalent to a 330ms frame exposure time. The recorded number of counts recorded per pixel is normalised to one second. The DCR is characterised with the sensor at 1.5V, 2.5V and 3.5V excess biases. Ten samples of revision 'AB' are used to characterise the DCR. The temperature of the IC packages was measured at 20°C mean using the temperature probes of an off-the-shelf electrical multi-meter. The quench voltage ('VQ') is biased at 0.7V producing a dead time >200ns across all excess bias conditions.

Figure 6.11(a) shows an example at 1.5V excess bias of a normalised DCR map with logarithmic scaling of one image sensor IC. Crosstalk is evident around the highest DCR SPADs with elevated count rates of the neighbouring pixels. Crosstalk cannot be quantitatively measured without the ability to enable and disable individual SPADs such as shown in [95]. Otherwise, as described in [140] the crosstalk may only be qualitatively assessed by cross-correlation analysis. Performing 2D cross correlation first on the image in Figure 6.11(a), and secondly on the same image data with all pixels randomly re-distributed, shows that that the original image has higher correlation, indicating that cross-talk is indeed present. Figures 6.11(b) and (c) display the output of the same sensor biased at 2.5V and 3.5V excess bias respectively. From a previous study [95], it is determined that well sharing increases electrical cross-talk, and the degree of cross-talk is proportional to the length of the shared dimension between SPAD devices. As pixels shrink, this boundary decreases and so scaling SPADs to lower pixel sizes should decrease cross-talk by this assumption.

Figure 6.12 shows the cumulative population graph versus count rate for all samples. Figure 6.13 plots the same data as a histogram. It is observed that the first populations increase substantially more for increasing excess bias, than the other populations in the histogram. It is assumed that the first population is due to thermal generation and the other populations are trap-assisted DCR, therefore it appears that during the experiment some degree of temperature increase has occurred on the die. Table 6.5 presents the median DCR and the three quartiles at three excess bias conditions. No compensation has been applied for the compression of count rate of high DCR SPADs due to the digitally oversampled binary image capture (detailed later in this chapter). In comparison, the median DCR of 'AA' chip revision is 312 CPS at 1.5V excess bias.

The DCR was improved in the 'AB' chip revision by process improvement (the details of which remain company confidential).

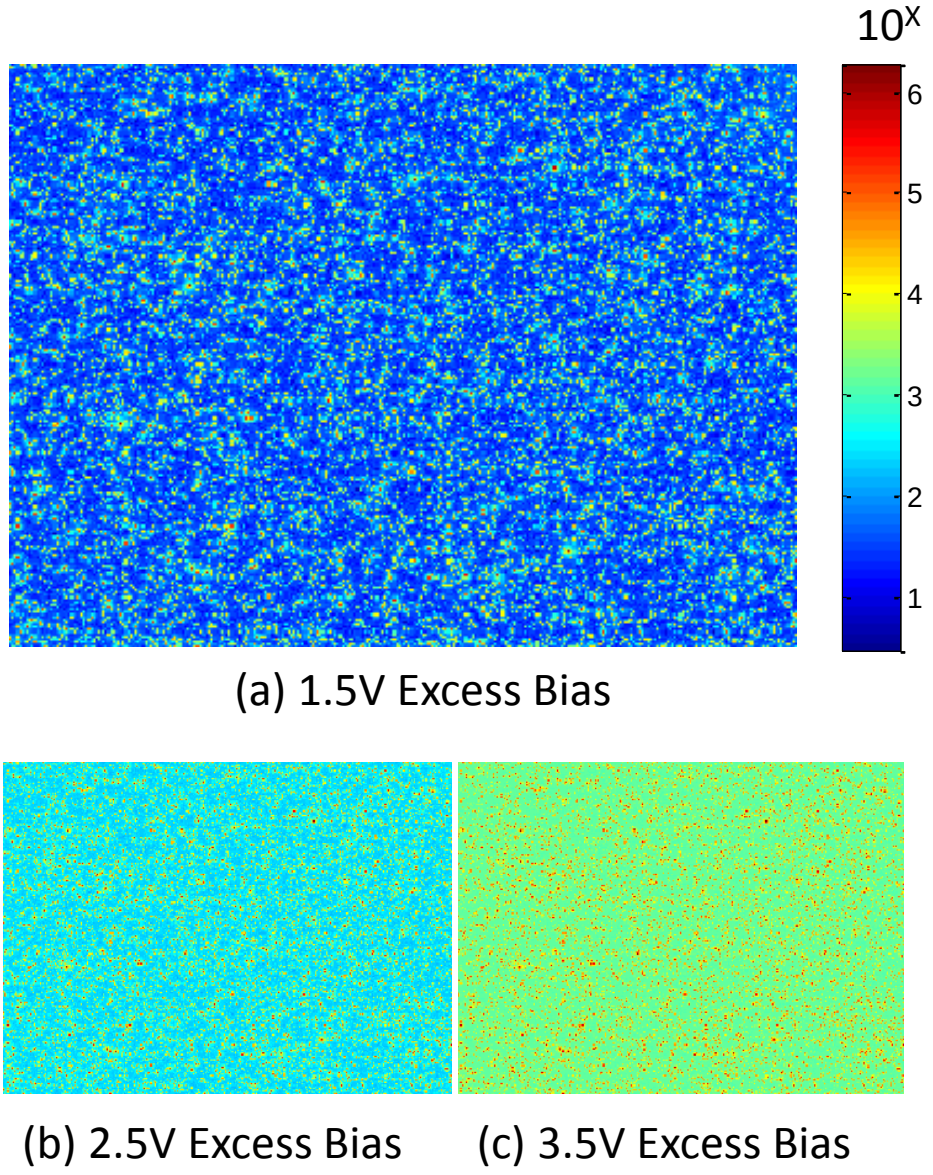


Figure 6.11. DCR map of the 'SPCIMAGER_AB', the scale is logarithmic in counts per second (CPS) at 20°C.

	Excess Bias (VHV Setting)		
	1.5V (14.8V)	2.5V (15.8V)	3.5V (16.8V)
Median (CPS)	55.8	346.3	1728.4
1st Quartile (CPS)	35.2	281.7	1502.4
Mean (CPS)	55.8	346.3	1728.4
3rd Quartile (CPS)	135.0	619.2	2459.1

Table 6.5. Table of median, first quartile, mean and third quartile DCR for three excess bias conditions for ten samples of 'SPCIMAGER_AB' at 20°C.

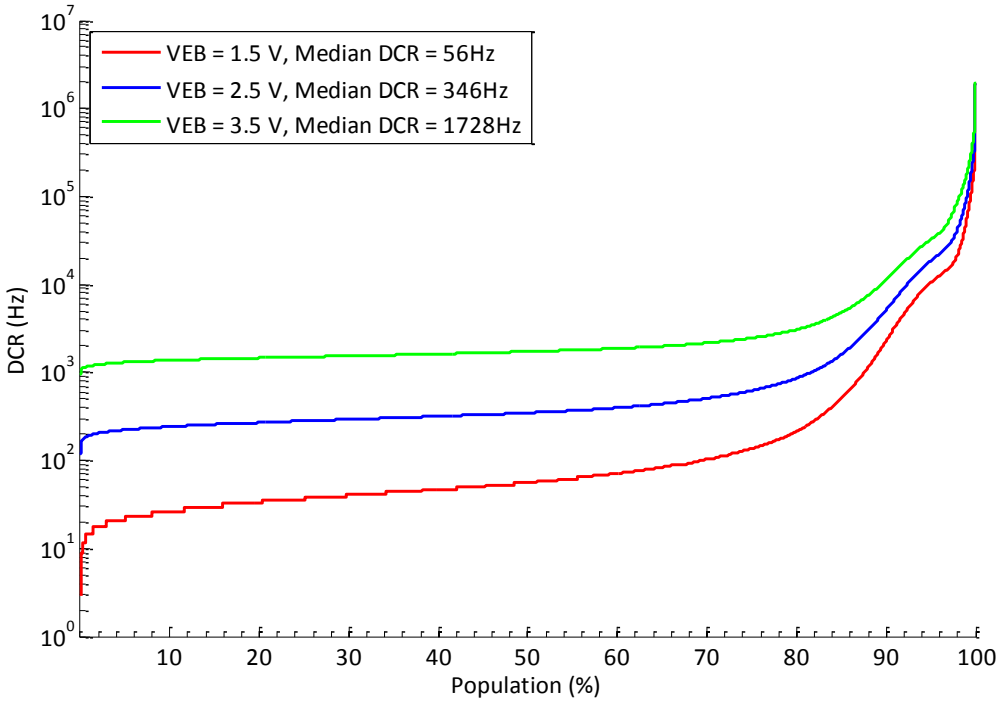


Figure 6.12. Cumulative plot of DCR versus percentage population for 10 samples of images sensor revision 'AB' at 20°C.

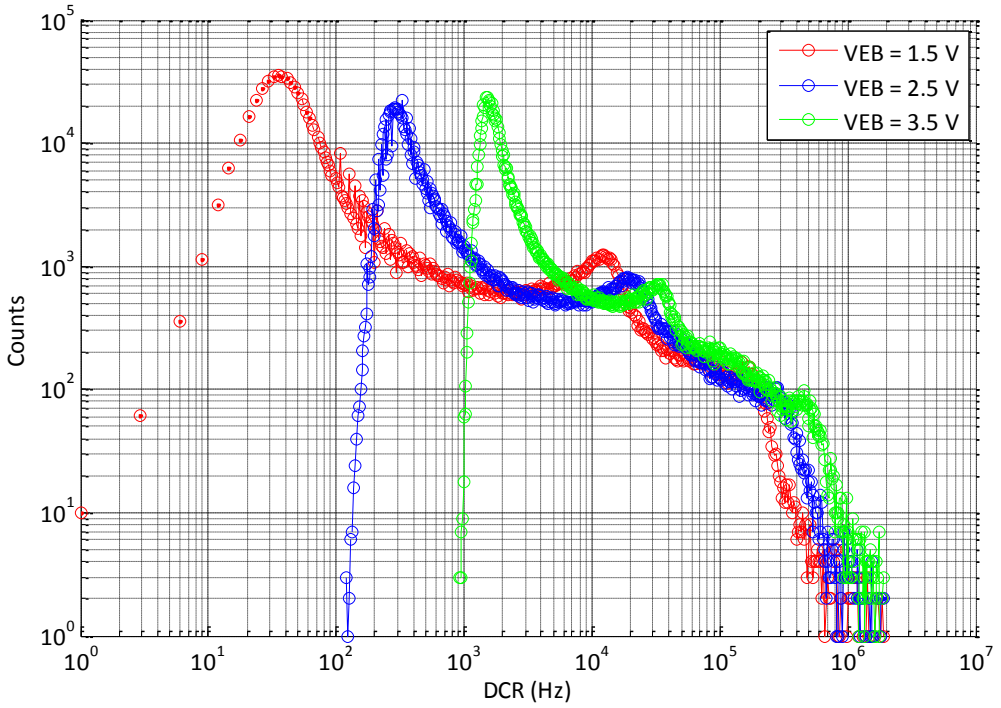


Figure 6.13. Histogram of DCR for ten samples of image sensor revision 'AB' at 20°C.

6.1.8. After-pulsing

After a SPAD has avalanched and recharged, a secondary avalanche may be triggered by the primary avalanche due to trapping or other device physical operations [68]. The effect of this, in single photon counting, is the mean number of recorded SPAD avalanches will be skewed, from the original photo-generated events, by the percentage of after-pulsing. Furthermore, in a TOF system this is a critical parameter as a secondary after-pulse, many nano-seconds after the first, will distort a TOF measurement.

Doctoral colleague Luca Parmesan conducted this experimental measurement and reported the results. A low-DCR SPAD was selected to avoid characterising the trap lifetimes in high DCR SPADs. A sweep of quench voltage and SPAD excess bias was performed to evaluate the change of after-pulsing over the SPAD operating region. The SPAD test IC was placed in dark room conditions at room temperature. The inter-arrival time of secondary SPAD pulses, after a primary SPAD pulse, was captured using a LeCroy 'WavePro' oscilloscope. For each iteration of the after-pulsing measurement (at each voltage setting), 500 after-pulses were captured for a variable number of primary pulses. Only a single SPAD device was measured due to time constraints, as the total measurement time exceeded four days.

Figure 6.14 (a) displays a typical histogram of inter-arrival times indicating the presence of after-pulsing with a characteristic exponential decay shape measured at $V_{HV} = 15V$, $V_Q = 2.2V$. For each setting of V_{HV} and V_Q, the after-pulsing is captured in a similar fashion and the recorded 500 secondary counts are divided against the number of primary SPAD pulses detected. Figure 6.14 (b) illustrates the measured after-pulsing at each voltage setting.

After-pulsing is nearly undetectable and may be considered negligible for $V_Q<1.8V$. As found in the previous dead-time measurements, unpredictable SPAD behaviour occurs for VQ>2.2V where after-pulsing at this PQPR bias condition significantly increases with VHV. After-pulsing using the method described above could not be accurately determined at $V_{HV} > 15.5V$, $V_Q \geq 2.2V$ as the pulse width has increased beyond the measurement window of the oscilloscope, whereas the dead time measurement in section 6.1.3 better illustrates this phenomenon. A suggested improvement to the measurement of after pulsing is conducting a real-time measurement capturing simultaneously both dead-time pulse width and after-pulse inter-arrival time.

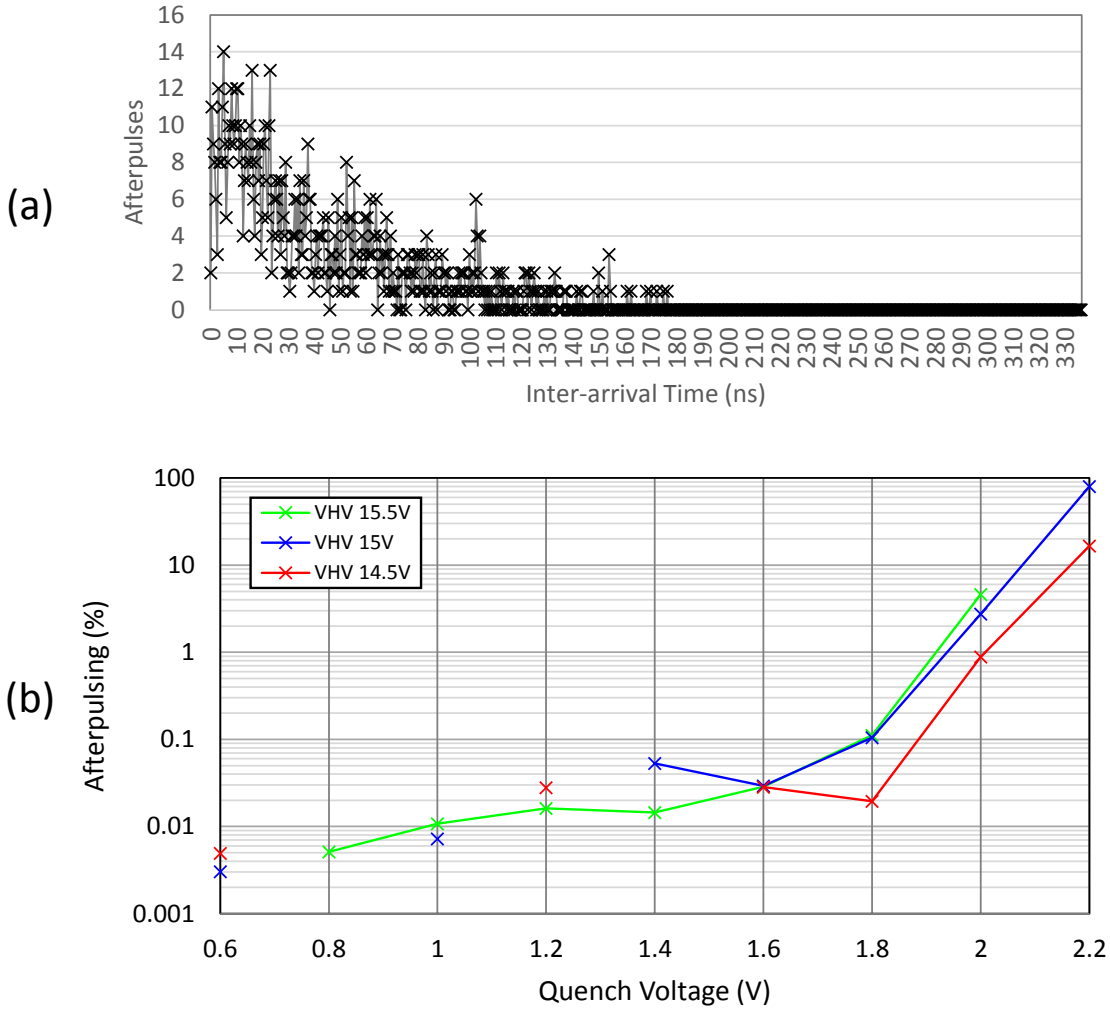


Figure 6.14. SPAD after pulsing for a single SPAD device: (a) shows a typical histogram of secondary SPAD pulse inter-arrival times indicating after-pulsing is present at V_{HV} = 15V, V_Q = 2.2V. (b) After-pulsing measured for a range of SPAD excess bias indicating a relationship to quench voltage.

6.2. Analogue Readout

The timing of control signals and evaluation of the noise contributions throughout the analogue readout chain are critical to ensuring the best performance is achieved with this single photon image sensor. Incomplete analogue settling will create artefacts, such as lag and crosstalk, and produce high temporal noise in the image. The timing and control of the sensor is briefly discussed and optimum settings or values are used in subsequent experiments. The readout noise sources are characterised to safeguard against noise sources corrupting or degrading the single photon counting performance of the sensor. This section details the analogue readout chain in reverse from the PCB power supplies, the off-chip ADC back through to the pixel.

6.2.1. PCB Power Supplies and Consumption

The image sensor PCB (*'PCB 1919A'*) schematics and package mappings are detailed in Appendix 3. There are two ground planes: quiet analogue ground 'AGND' and FPGA digital ground 'DGND'. These two planes are connected at a single point by a jumper to maintain the potential. The 5V PCB power input is fed from an off the shelf 240V AC to 5V DC switched power supply adaptor. The FPGA has its own on-chip regulators to generate its supplies for core logic and I/O banks. The FPGA ground is connected to the DGND plane on the PCB, which connects onto the chip in the I/O ring for digital I/O. Table 6.6 lists the sensor supply names versus the PCB supply and ground names for each of the blocks in the analogue readout chain. The DGND PCB ground will be the noisiest portion of the PCB and so is marked in red. Each of the PCB supplies to the sensor are generated by a separate DAC-controlled Linear Technologies regulator ('LT3080') powered by the 5V external supply.

	Block	Sensor Supply	Sensor Ground	PCB Supply	PCB Ground
Sensor	Pixel	VRT (V2V7)	AGND	V2	AGND
	RST Row Driver	V3V6	AGND	V3	AGND
	EN/DIS/RD Row Driver	V3V3	AGND	V1	AGND
	CDS Column Buffer				
	Dual Op-Amp's	VDDOPAMP	VSSOPAMP	V4	AGND
	Digital Logic	VDD	GND	1V2	AGND
	Digital I/O	VDDE3V3	GNDE	VDDE	DGND
PCB	ADC Ana' Front End	N/A	N/A	VDDRAMP	AGND
	ADC Digital Output	N/A	N/A		DGND
	FPGA	N/A	N/A	5V	DGND

Table 6.6. Sensor and PCB1919A power supplies and ground connections. The 'DGND' PCB ground plane is marked in red.

The power consumption of the 'SPCIMAGER_AA' was measured in both analogue and digital readout modes with a 5MHz clock under continuous image capture. Table 6.7 displays the measured current and calculated power consumption of each of the sensor supplies. In total the analogue readout consumes 6.1mW and the digital readout 5.7mW during image capture. The SPAD cathode 'VHV' supply is characterised separately in section 6.1.4.

	Analogue Readout		Digital Readout – 5.12kFPS	
Supply Name	**Current (µA)**	**Power (µW)**	**Current (µA)**	**Power (µW)**
V3V3 (3.3V)	2	6.6	562	1854.6
V3V6 (3.6V)	1	3.6	5	18.0
VRT (2.7V)	264	712.8	264	712.8
VDD (1.2V)	164	196.8	2280	2736.0
VDDOPAMP (3.3V)	1540	5082.0	0	0.0
VDDE (3.3V)	37.4	123.4	120	396.0
Total Power		6125.2		5717.4

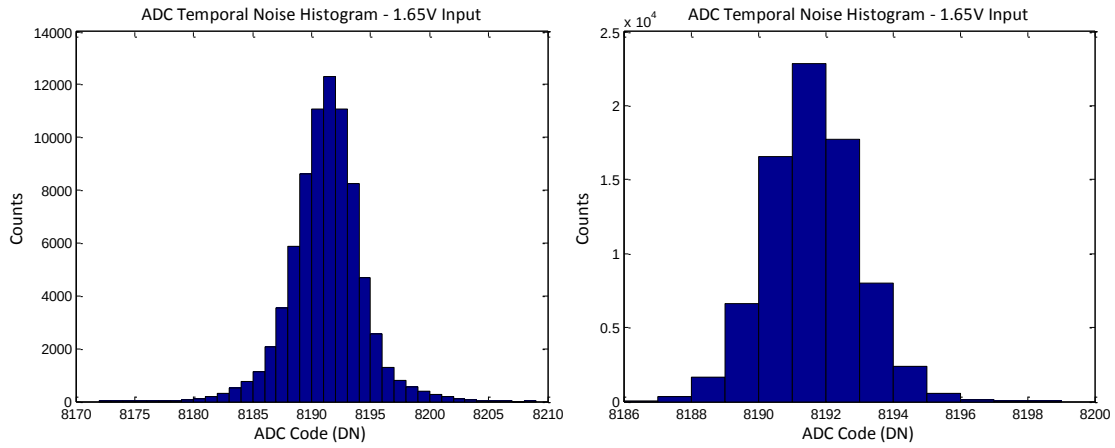
Table 6.7. Image sensor power consumption.

6.2.2. ADC and Op-Amp Temporal Noise

The temporal and cyclic noise contributions of the 14b pipelined Analog Devices ADC (part number *'AD ADC14L020'*) and output op-amps are common to all pixel signals and images captured using the analogue readout. As detailed in Table 6.6, the ADC has its own regulated power supply and has two grounds on-chip: 'AGND' for the front end and 'DGND' for the pipelined digital output. The temporal noise of the ADC is evaluated by shorting the inputs of the ADC together and applying a bias on the positive and negative inputs. To produce realistic experimental conditions, the sensor is used to drive the ADC inputs from the sensor op-amps and an image frame is captured, with all the sensor control signals operating. The SPADs are biased below breakdown, the exposure control does not activate the time gate and the pixel supply is pulled to ground in order to provide a constant input to the ADC.

Correlated multiple sampling (CMS) is a common CIS technique which increases the number of ADC samples in order to reduce the temporal noise contribution of the readout and converter electronics. Capturing 'N' samples reduces the temporal noise contribution (of the ADC and op-amps) by a factor of $\sqrt{N}$, but increases the readout time and decreases frame rate. A trade-off is made between the number of samples, noise reduction and frame rate. Consequently single and quadruple sampling are investigated. The temporal noise contributions of the ADC and output op-amps are measured with one and four samples. Figure 6.15 shows the two histograms showing a notable decrease in the distribution width using four ADC samples. Table 6.8 presents the calculated RMS values of the noise distributions, where the ADC noise is halved with four sample CMS. Four ADC samples are used in the remainder of the experimental work in this thesis. However, this is not true CMS, as a double sampling circuit is still employed to perform the initial sample of the column bus.

The ADC multiple sampling is only performed on the column buffer readout. True CMS would be performed with column parallel ADCs. The CDS terminology is therefore still applied in the rest of this chapter.



(a) 1 ADC Sample (b) 4 ADC Samples

Figure 6.15. Histograms of ADC noise with (a) 1 ADC sample and with (b) 4 ADC sample averaging.

No. of Averaged Samples	RMS value in ADC Codes	RMS value in Voltage
1	2.83 DN	346.1μV
4	1.37 DN	167.3μV

Table 6.8. One standard deviation of ADC noise in ADC codes (DN) and voltage (V).

6.2.3. Op-Amp Settling

The settling of the analogue output signals is confirmed by means of monitoring the bus voltages on an oscilloscope. Figure 6.16 demonstrates an annotated screenshot from the LeCroy 'WavePro' oscilloscope, the top yellow 'Op-Amp Differential' waveform is the computed differential (of the blue 'reset' waveform minus the brown 'signal' waveform) as seen at the ADC input. The op-amps have completely settled during the ADC sampling periods marked in pink. The oscillations on the op-amp signals are a coupling of the ADC clock signal either due to cross-coupling on the PCB or emanating from the ADC switched capacitor front end. For rising voltages, the op-amps output stage PMOS are evidently slew-rate limited.

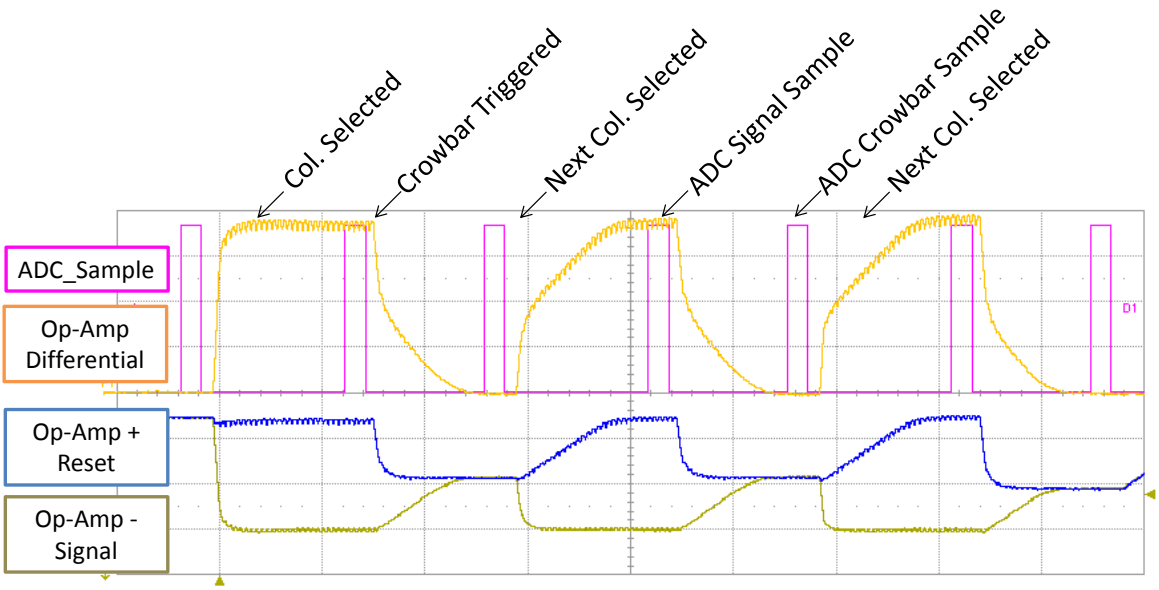
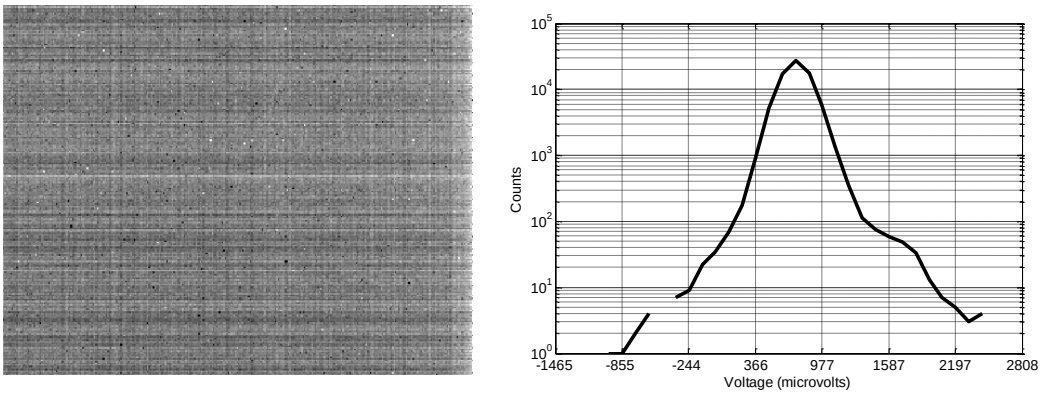


Figure 6.16. Annotated screenshot from LeCroy oscilloscope showing op-amp settling with full-scale inputs. The ADC sampling periods are marked in pink against the output op-amp differential signal.

6.2.4. Dark FPN

The dark FPN of the image sensor is evaluated through a rolling sensor reset timing mode which resets each pixel row before the CDS readout (firmware '*EXPOSURE_MODE*' register set to 8). The sequence consists of pixel reset then the standard delta reset CDS sequence of read signal, pixel row reset, read reset. The output image from this sequence will have temporal noise sources (thermal, flicker, charge injection, etc.) from the readout chain and dark FPN. To mitigate the contribution of temporal noise sources, 100 images are averaged. Figure 6.17(a) is the average image of the black level showing both visible horizontal and vertical FPN. Figure 6.17(b) is a histogram of this averaged image. The horizontal, vertical and pixel-to-pixel FPN components (HFPN, VFPN and PPFPN respectively) are calculated and shown in Table 6.9.

As seen in the non-zero values in the table, the CDS scheme employed on the sensor does not fully remove dark FPN. An algorithm is written in the MATLAB sensor control software to automatically learn the dark FPN image as the sensor is switched on, before analogue SPC image capture. This improves the image capture by suppressing the dark FPN down to negligible levels, leaving the temporal noise sources.



(a) Dark FPN Image (b) Dark FPN Histogram

Figure 6.17 (a) Dark FPN averaged image. (b) Histogram of black level with logarithmic axis.

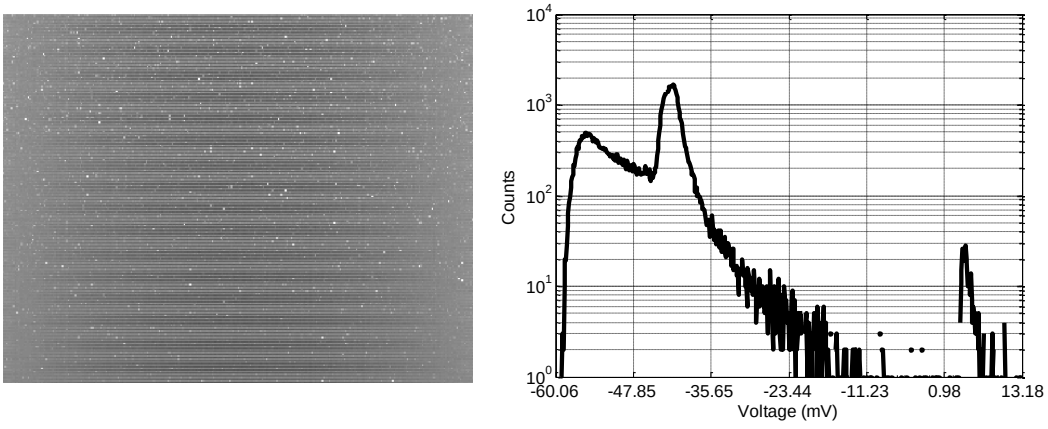
	HFPN	VFPN	PPFPN
Std. Dev. (ADC Codes)	0.746 DN	0.660 DN	0.811 DN
Std. Dev. (Voltage)	91.1μV	80.6 μV	99.0 μV

Table 6.9. Dark FPN horizontal, vertical and pixel to pixel components.

6.2.5. Global versus Rolling Reset

There are two distinct options for the initial full reset of the pixel array before each exposure, 'rolling' row by row sequentially or 'global' with all pixels simultaneously reset. The rolling reset is activated for twenty clock cycles per row (1µs on a 20MHz clock) or 4800 clock cycles total (240µs). Whereas, the global reset is activated for a total of forty clock cycles (2µs on a 20MHz clock) which has the advantage of permitting a faster frame rate. In the following experiments, the SPADs are held off below breakdown, the analogue counter is biased with a negative V_{GS} (V_S = 0.2V, V_G = 0V) to reduce leakage through the counter, and the sensor is in darkness. The initial global reset is first evaluated by performing a blank 'zero-second' exposure, followed by the rolling delta reset CDS readout. The zero-second exposure timing ensures the time-gate disable is always high and the time gate is not triggered. To mitigate supply noise, a ten sample average image of the global reset black level is taken and is shown in Figure 6.18(a). Here the dark FPN black level learning procedure, discussed previously, is not employed. A histogram of the global shutter dark image is shown in Figure 6.18(b). An interleaved pattern is evident in the image and seen in the two merged peaks of the left of the histogram.

To evaluate this interleaved pattern, Figure 6.19 splits the global reset dark image into two, the first consisting of odd rows and the second of even rows. The odd row image has a distinct horizontal shading pattern due to the V_{RT} IR drop, whereas the even row image has a uniform distribution across the image. On investigation, the pixel layout in both 'AA' and 'AB' sensor revisions was found to be missing a metal via to the vertical power connection on the odd rows.



(a) GS Initial Reset Dark Image (b) GS Initial Reset Dark Histogram

Figure 6.18 (a) Image of black level captured under global initial reset. (b) Histogram of the image.

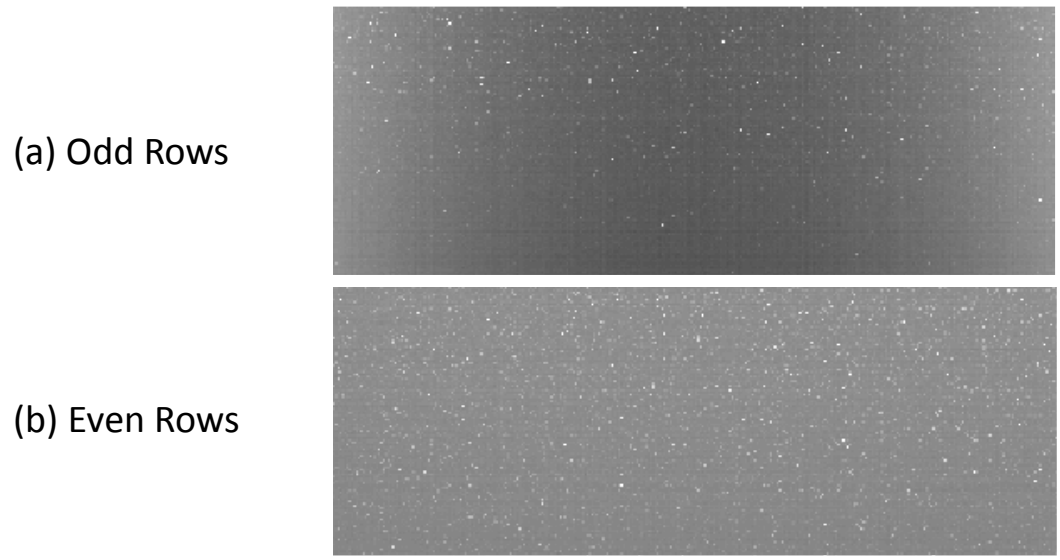
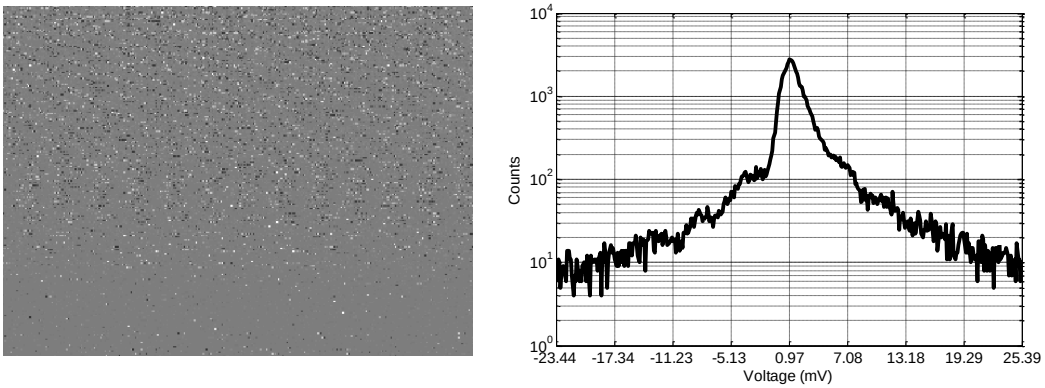


Figure 6.19. Global initial reset image split into (a) odd rows and (b) even rows. The odd rows image has a power supply IR drop evident in the horizontal shading. The same contrast is used in both images.



(a) RS Initial Reset Dark Image (b) RS Initial Reset Dark Histogram

Figure 6.20. Rolling initial reset, rolling CDS reset. (a) Image of black level. (b) Histogram of image.

The rolling initial reset is evaluated in the same manner by performing a zero-second exposure. Figure 6.20(a) is a ten sample averaged image of the rolling reset black level and (b) is the histogram. The row pattern is not evident indicating the offset, due to the IR-drop, has been corrected for in the CDS operation.

On the one hand, the global reset produces a faster frame rate, yet as the CDS readout performs a rolling reset, the cancellations of reset noise and the offset due to the IR drop are not corrected for. The high variation or shading in black level (due to IR droop on alternate rows) with the global reset is not acceptable for the analogue readout mode and so rolling initial reset is implemented for the remainder of the analogue counting

experimental work and results taken within this thesis. However, in the digital readout mode, no CDS is performed and mV variations in the reset level are tolerable, whilst ensuring the binary decision threshold is clear of the reset level. Moreover, the digital mode requires the fastest possible sensor operation forcing the minimum time possible for the initial reset before each image field, and so global reset is implemented.

6.2.6. Temporal Noise

The temporal noise from a single image through the analogue read-out chain, after post capture FPN correction (100 frame averaged dark-frame subtraction), is evaluated at 916μV RMS. This figure represents row and column temporal noise, kT/C contributions in the readout from column capacitors, column source follower and op-amp temporal noise and ADC cyclic and temporal noise. This value may then be input-referred to evaluate the read noise of the analogue single photon counting mode. Due to the limitations of the control of the sensor, the individual noise contributions of each of the blocks cannot be isolated. The contributions that can be isolated are:

- The op-amp and ADC temporal noise contribution, after 4 sample CMS averaging, is 167.3μV RMS.
- The column sample and hold capacitors and the two column buffer readout paths are not linked until the ADC front end, therefore both reset and signal capacitors kT/C contributions both add independently. The kT/C contribution of each of the column capacitors (176.2fF) is calculated to be 151μV RMS assuming 20°C.
- The delta reset CDS timing scheme employed (as opposed to true CDS timing) will not cancel the kT/C noise of the pixel capacitor, although it will be high-pass filtered by the CDS operation reducing source follower flicker noise (where the high pass filtering bandwidth is proportional to the reset period and the time difference between column samples [43]). The lumped pixel capacitor value in layout extraction is 15.1fF, which provides a calculated 518μV RMS kT/C noise contribution without taking the bandwidth filtering of the pixel kT/C noise into account.

Subtracting all these three contributions in quadrature leaves a contribution of 704μV RMS from the row supply temporal noise, pixel source follower (RTS and flicker) and column readout temporal supply noise. This high value indicates that the on-chip supply decoupling may be improved in future revisions of this sensor and generally highlights the importance of considering noise performance and reduction in an image sensor design process.

6.3. Analogue Single Photon Counting

The sensitivity or step size of the analogue single photon counting is measured against SPAD excess bias, SPAD quench voltage, and counter gate and source biases. Due to leakage the minimum source (V_S) bias must be maintained at greater than 0.15V. Like the single photon counting images and histograms in the previous section, the acquired data are captured with a single frame (no averaging) and dark FPN correction is applied.

6.3.1. Counter Sensitivity to VS Bias and Excess Bias

To evaluate the counter sensitivity to both counter source bias and SPAD excess bias, the sensitivity equation (Eq.4.5) is updated to match the image sensor pixel circuit diagram (with M7 becoming M5):

$$\Delta V_C = \left(\frac{C_P}{C_T}\right) . (V_{IN} - V_S - V_{TM5})$$

(Eq.6.3)

First the relationship of source bias to sensitivity is analysed to evaluate the capacitor ratio. The V_G bias was set at 0.5V above VS, for each of the captures. The linear relationship of CTA step size to V_S bias is shown in Figure 6.21. Below 8.5mV per SPAD event sensitivity it is difficult to obtain discrete SPC peaks.

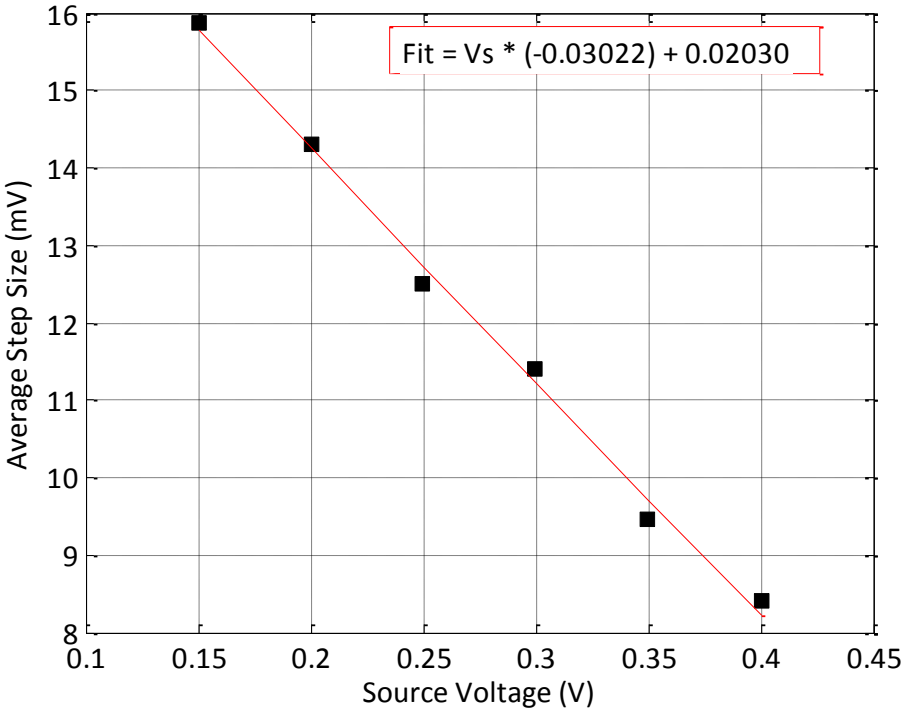


Figure 6.21. Mean step size versus applied V_S bias voltage. A linear fitted line is plotted for comparison.

A linear fit is calculated and is plotted alongside, and is in the form:

$$\Delta V_C = \left(\frac{C_P}{C_T}\right)(-V_S) + \left(\frac{C_P}{C_T}\right).(V_{IN} - V_{TM5})$$

(Eq. 6.4)

Secondly, the effect of increased SPAD excess bias on the counter sensitivity is shown in Figure 6.22 with a linear fit similarly plotted beside. The clamping effect of the inline time gate switch is evident for the SPAD cathode V_{HV} voltage greater than 17V. The maximum recorded sensitivity is 17.4mV per SPAD event at 18.5V SPAD cathode voltage. A linear fit is calculated and plotted for $V_{HV} < 17V$, and is in the similar form:

$$\Delta V_C = \left(\frac{C_P}{C_T}\right)(V_{IN}) - \left(\frac{C_P}{C_T}\right).(V_S + V_{TM5})$$

(Eq. 6.5)

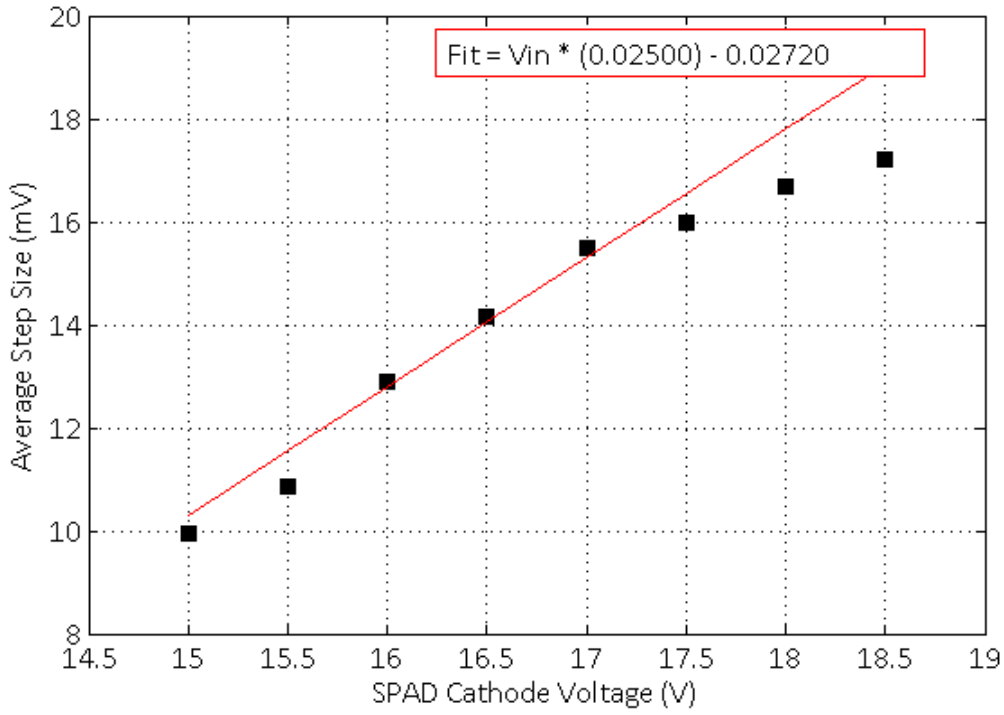


Figure 6.22. Mean step size versus SPAD cathode voltage. A linear line is plotted for comparison.

From the two linear fit equations, the capacitor ratio can be computed. Table 6.10 compares the capacitor ratio of the test pixel structure in Chapter 4, to the image sensor described here in Chapter 6. The capacitor ratio, of the imager pixel, is approximately twice that of the test structure. Reviewing the layout of the two pixels, the area of the N+ diffusion parasitic capacitance is 2.8 times smaller in the image sensor although the gate overlap capacitances remains the same.

However, as the counter operates in the voltage domain it can achieve a similar full well to the test array by adjusting either VS bias or the SPAD excess bias to compensate.

	Chapter 4 Test Pixel Array		**SPCIMAGER AB**	
	Experimental Data Fit Pulse Test– VS Bias	**Experimental Data Fit SPC – VS Bias**	**Experimental Data Fit SPC – VS Bias**	**Experimental Data Fit SPC – Excess Bias**
Capacitor Ratio (C_P / C_T)	0.01309	0.01383	0.03022	0.02500

Table 6.10. Comparison of capacitor ratio linear fit parameters from Chapter 4 test pixel array, and image sensor.

6.3.2. Counter Sensitivity to VG Bias

Figure 6.23 demonstrates the effect of increasing the gate-source voltage of the M8 CTA discharge transistor on the average step size (calculated for six steps) and standard deviation of the first voltage peak. The source is maintained at 0.2V and the SPAD is biased with 16.5V cathode voltage and 1V quench bias. These graphs highlight the transition region between CTA mode with 3.3mV RMS step size (the leftmost two samples) and SCS mode with > 50mV step size (the samples on the right). After $V_G > 0.9V$ it becomes increasingly difficult to measure the step size as there is a large spread with no discernible peak, and after VG>1.2V the output becomes a single step indicating a full discharge of the pixel capacitor below the threshold of the in-pixel source follower. The pixel functionality becoming the desired SPAD-activated dynamic memory discharge operation, for increasing V_G bias voltage above 1.2V although this is not shown below. This is in line with the previously measured results in Chapter 4.

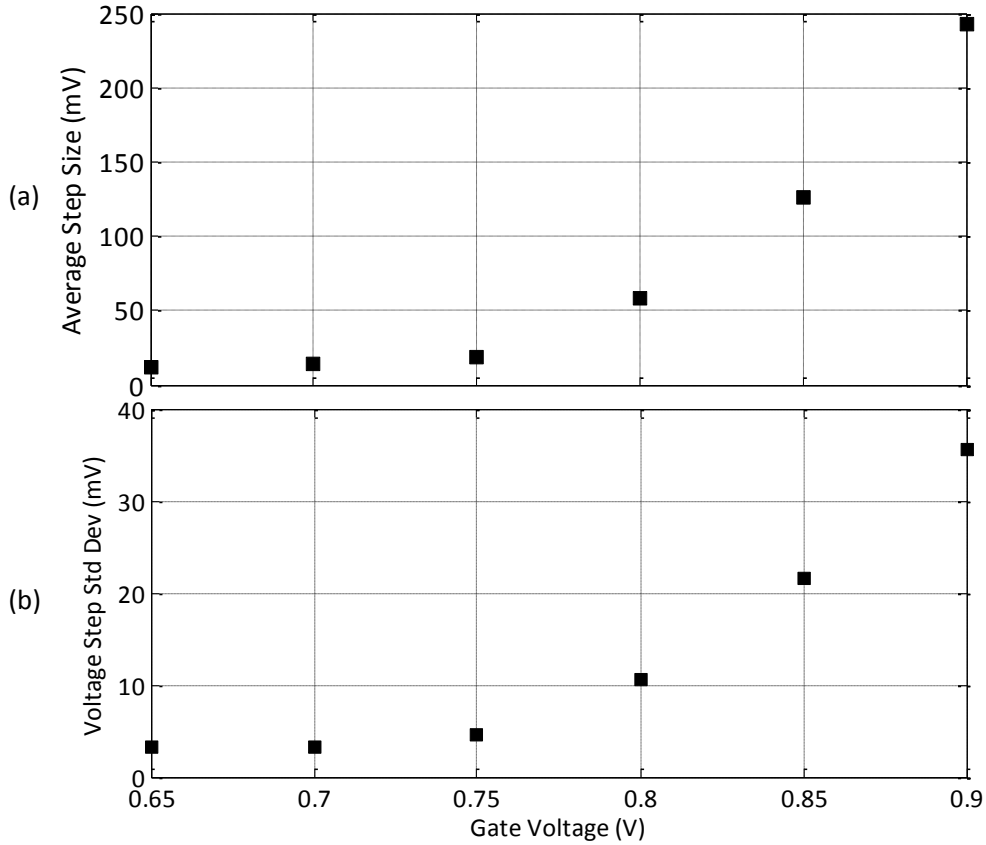


Figure 6.23. Effect of increasing gate voltage on (a) the average step size and (b) on one standard deviation of the first step.

6.3.3. Counter Sensitivity to Quench Voltage

According to the first order theoretical CTA equation, the quench voltage should have no bearing on the counter step size. However, as shown in Figure 6.24 the counter step size has a non-linear dependence on the quench voltage. The dependence has the same shape as the quench resistance and dead time characterisation. There are two hypotheses for the interaction of quench voltage to step size. Either the increased quench voltage reduces the peak voltage on the CTA input node (effectively reducing the excess bias seen by the counter circuit), or this indicates that the SPAD dead time does affect the CTA sensitivity. The longer dead time (at lower quench voltage) may increase leakage.

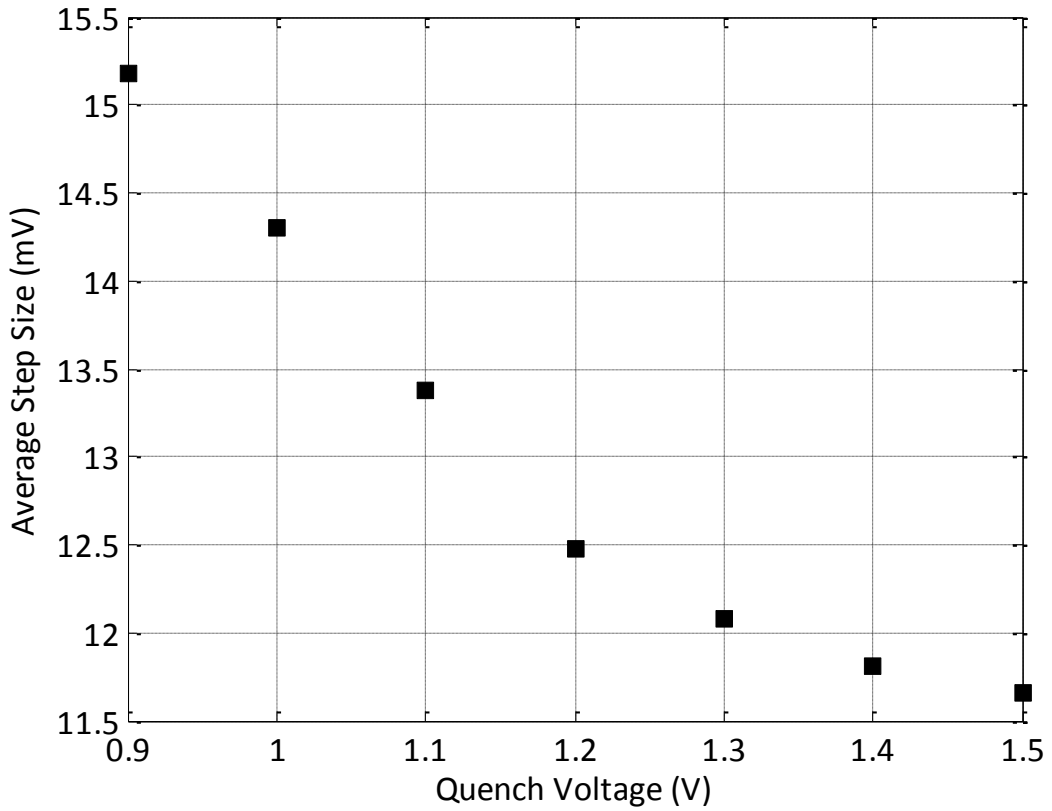


Figure 6.24. Mean step size versus applied quench voltage.

6.3.4. Single Photon Counting

SPC is achieved using the sensor analogue readout (with FPN frame subtraction) and the pixel biased in CTA mode. Figure 6.25 illustrates a typical histogram of the whole sensor output for a single capture under uniform LED lighting operated with constant current. A sensitivity of 14.2mV per SPAD event is recorded and marked on the figure. The following bias conditions were used in the experiment:

- VHV = 18.5V
- VQ = 1.0V
- VS = 0.15V
- VG = 0.7V

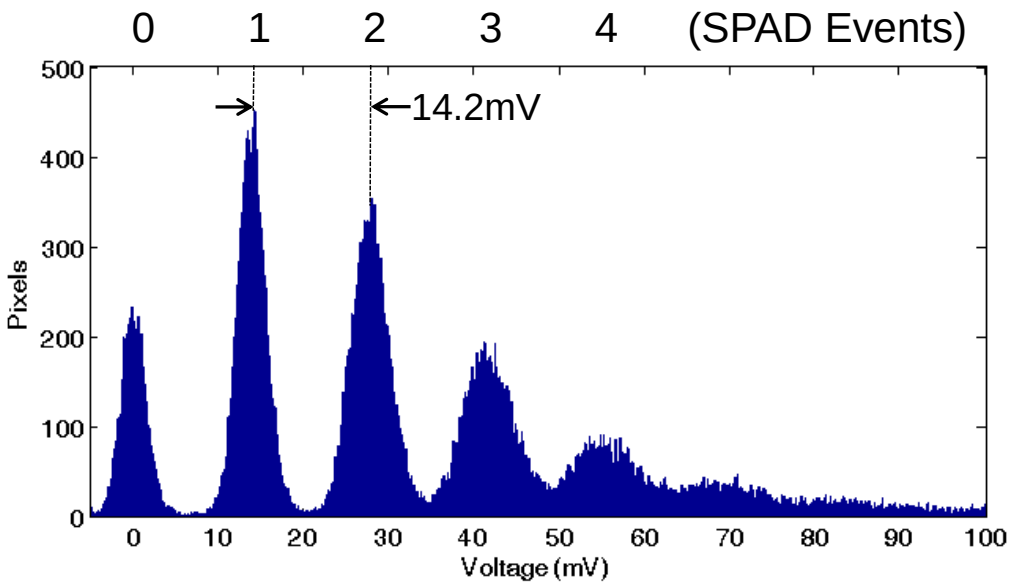
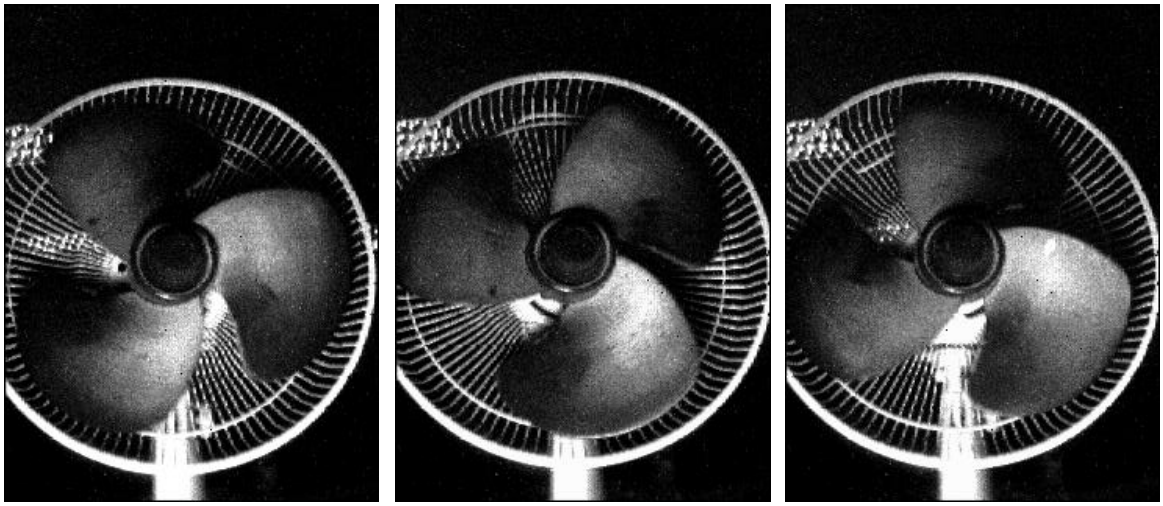


Figure 6.25. Whole sensor output histogram for 14.2mV sensitivity for single photon counting under constant LED lighting.

6.3.5. High Speed Global Shutter Image Capture

The image sensor is configured for photon counting using the analogue readout. Figure 6.26(a) below illustrates three global shutter images taken of a fast rotating fan with 300µs exposure in portrait orientation with an F#2.0 lens. Office lighting and a desk lamp are illumination sources. A post-process threshold median filter has been applied for high noise pixels. Figure 6.26(b) is a comparative image of the same scene captured with a 38 mega pixel (MP) CIS showing the rolling shutter distortion of the blades. This is a qualitative demonstration of the capability of the image sensor to capture fast occurring (sub-ms) optical phenomena with high sensitivity. Although not the focus of this research, the sensor in this mode is suitable for scientific, automotive, industrial and machine vision imaging applications.

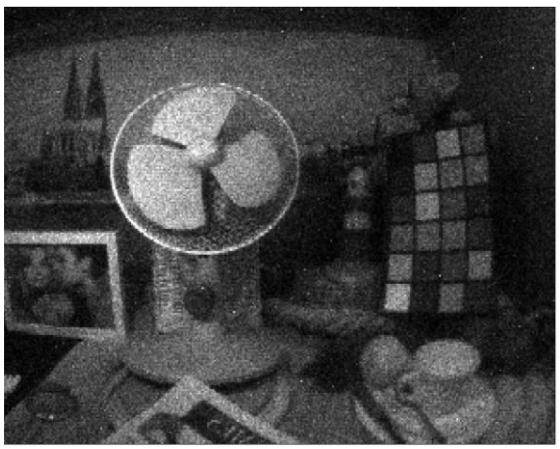
The two images in Figure 6.27 are taken in dark room conditions with a moving fan using a calibrated light environment. The contrast of the two images is scaled independently and the image on the right has had a bright pixel threshold median filter applied to suppress higher DCR or 'defect' pixels. The image on the left indicates that low-light imaging is possible with the sensor. However, the dark noise of the SPAD-based sensors is orders of magnitude higher than conventional CIS which prevents imaging for millisecond long exposures. Yet as demonstrated in the right image in the figure, micro-second short exposures can be captured in low light. This indicates that high-speed imaging at low-light is viable despite the high dark noise.



(a) 300µs global shutter exposures of rotating fan.

(b) Same scene taken with a 38MP CIS.

Figure 6.26. (a) Three global shutter images captured using the 'SPCIMAGER_AA' (VS=0.2V, VG= 0.7V) with 300µs exposure time. (b) The same scene was captured using a 38MP rolling shutter CIS.



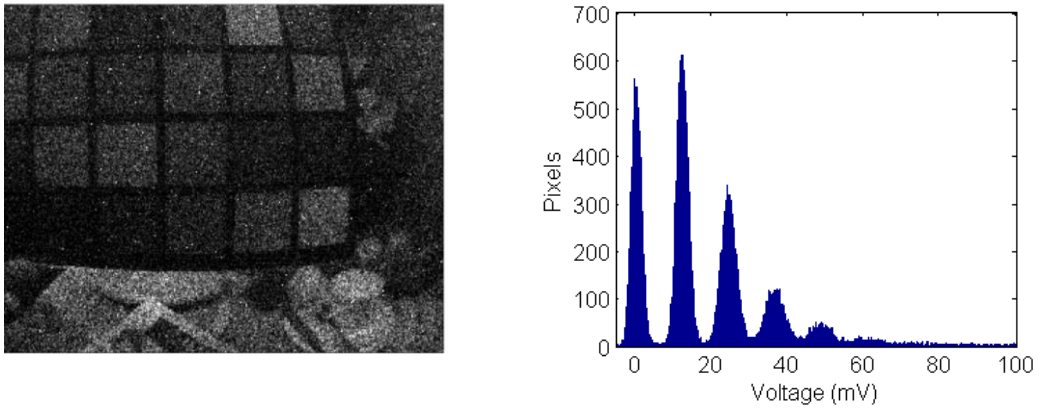
(a) 5µs exposure, 400 Lux

(b) 5µs exposure, 5 Lux

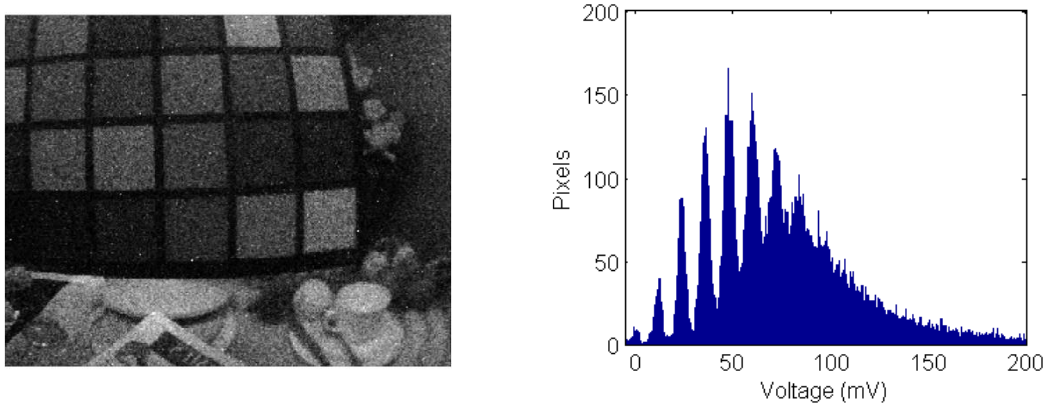
Figure 6.27. Global shutter images captured at 400 and 5 lux with a moving fan.

6.3.6. Input Referred Noise Measurement and Modelling

Two examples of SPC imaging are shown in Figure 6.28, where the contrasts, of the two images, are scaled independently. The images are captured with a fish-eye F#2.0 lens causing a spatial distortion of the calibration chart. The histogram for each image is also shown alongside.



(a) 5µs exposure, 100 Lux



(b) 20µs exposure, 100 Lux

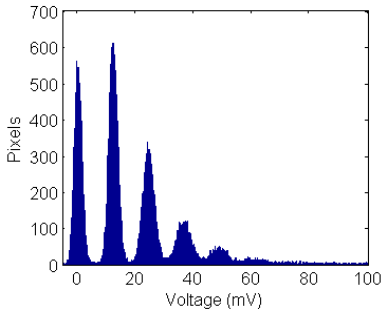
Figure 6.28. Single photon counting images taken with two exposures times. The histogram of each image is shown to the right.

In Figure 6.28(b) the accumulated noise in the right half of the histogram is clearly discernible. The accumulated noise modelling performed in Section 4.4.6, is repeated to evaluate this effect. In the previous test array, in order to measure the accumulation of noise per step, the increasing width was recorded of each peak in a SPC histogram of an accumulated exposure. Unfortunately repeating this measurement with the image sensor achieves the capture of the first peaks (as seen in Figure 6.28(a)) but only the widths of the first few peaks can be accurately determined, after which the accumulated noise engulfs the discrete peaks.

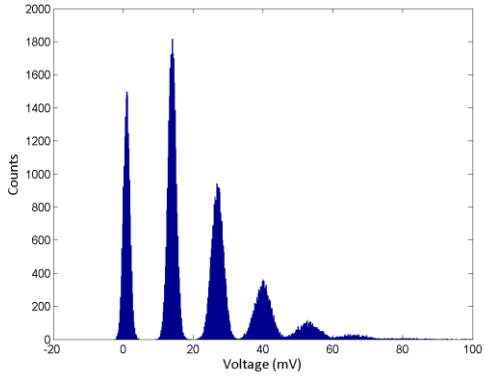
However, performing the accumulated noise modelling allows a value to be determined through an iterative process. The following parameters are used in the modelling:

- An ideal Poissonian distribution of 76,800 photons.
- A counter step size of 10mV.
- Read noise of 916µV RMS, (equivalent to input referred read noise of 0.09e-).
- 700µV RMS accumulated noise per step.

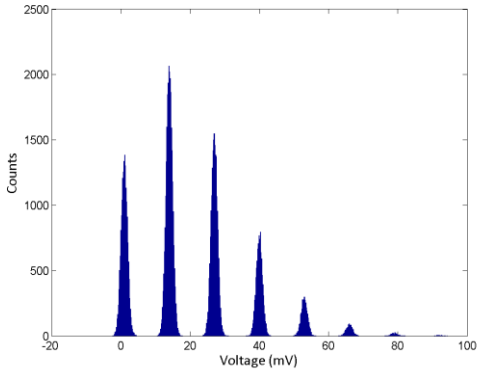
Figure 6.29(b) embodies the closest found fit, demonstrating the modelled simulation data (with read noise and accumulated noise) against the SPC histogram captured and shown in the previous Figure 6.28(a) and repeated for ease of comparison in Figure 6.29(a). To contrast to this accumulated noise histogram, Figure 6.29(c) illustrates the modelling of the same ideal Poissonian distribution (1.5 photons mean) with solely read noise, representing the histogram that could be expected for a true-electron counting system with 0.09e- input referred read noise.



(a) Measured 20µs exposure, 100 Lux



(b) Modelled Histogram with Read Noise and Accumulated Noise

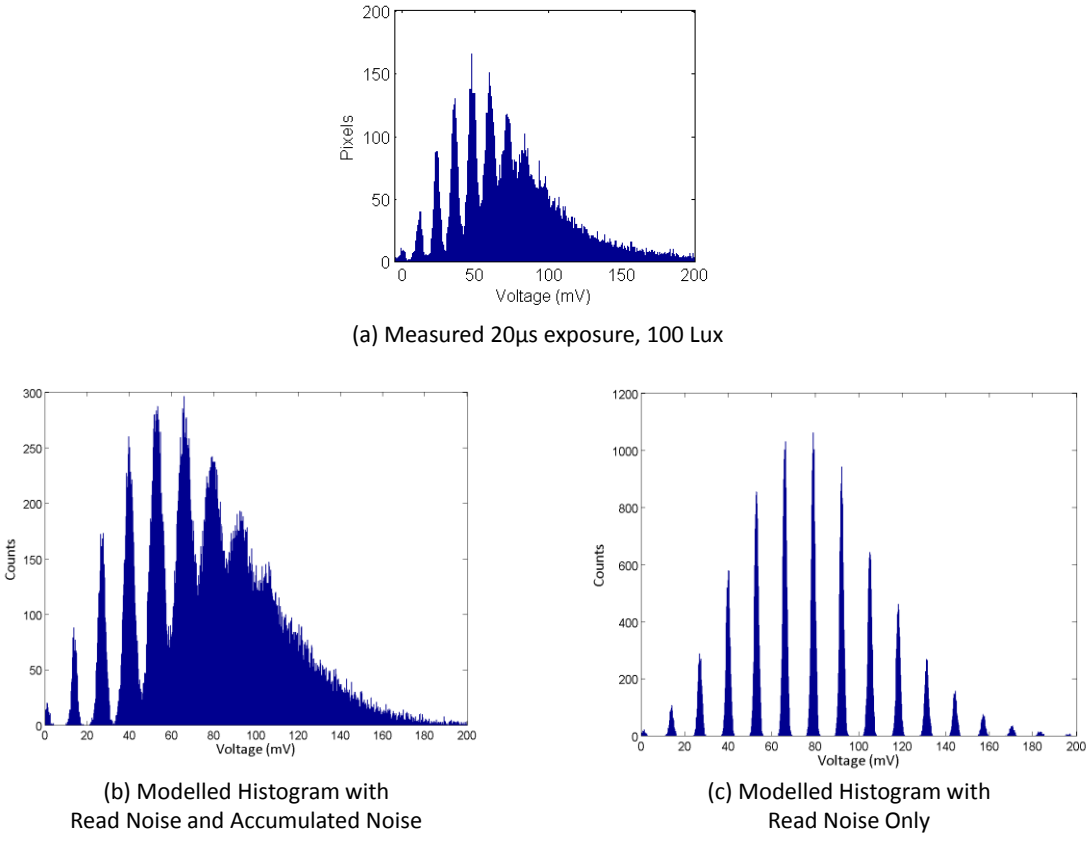


(c) Modelled Histogram with Read Noise Only

Figure 6.29. Accumulated noise modelling for mean of 1.5 photons. (a) Measured histogram from Figure 6.28. (b) Modelled histogram with read noise and accumulated noise. (c) Modelled histogram of true electron counting with only read noise.

Figure 6.30 repeats the exercise demonstrating the accuracy of the simulation to show the noise accumulation in the right half of the histogram. Both figures illustrate the closest found fit at 700µV RMS accumulated noise per step. This is an increase of eight times over the previous analogue counter test structure presented in Chapter 4, with an increase of 9 to 76800 pixels.

Two comments are merited on the matching of the modelling to the experimental data. Firstly, the CTA distortion, due to the imperfect discharge of the parasitic, is not modelled as it cannot be determined by this method. Secondly, the models both have higher recorded counts in each peak than the experimental data. In the model, the mean number of photons is assumed constant, whereas the measured data is for an imaged scene with non-uniform illumination across the array.



(a) Measured 20µs exposure, 100 Lux

(b) Modelled Histogram with Read Noise and Accumulated Noise

(c) Modelled Histogram with Read Noise Only

Figure 6.30. Accumulated noise modelling for mean of 6 photons. (a) Measured histogram from Figure 6.28. (b) Modelled histogram with read noise and accumulated noise. (c) Modelled histogram of true electron counting with only read noise.

For analogue single photon counting structures which contribute additive noise, it can be determined that the input-referred noise is not constant. Evaluating both the image sensor and the test structure from Chapter 4, Figure 6.31 plots the modelled input referred noise as a function of the number of SPAD-triggered analogue counter steps for a sensitivity of 14.2mV per SPAD event (as before using one SPAD event as the equivalent of a photo-generated electron). In the test structure, true CDS timing was applied which cancels the kT/C noise of the pixel capacitor, however, this is not the case with the imager with delta reset sampling. This kT/C noise, plus an increase in temporal noise, is observed as the difference in offset of the two curves.

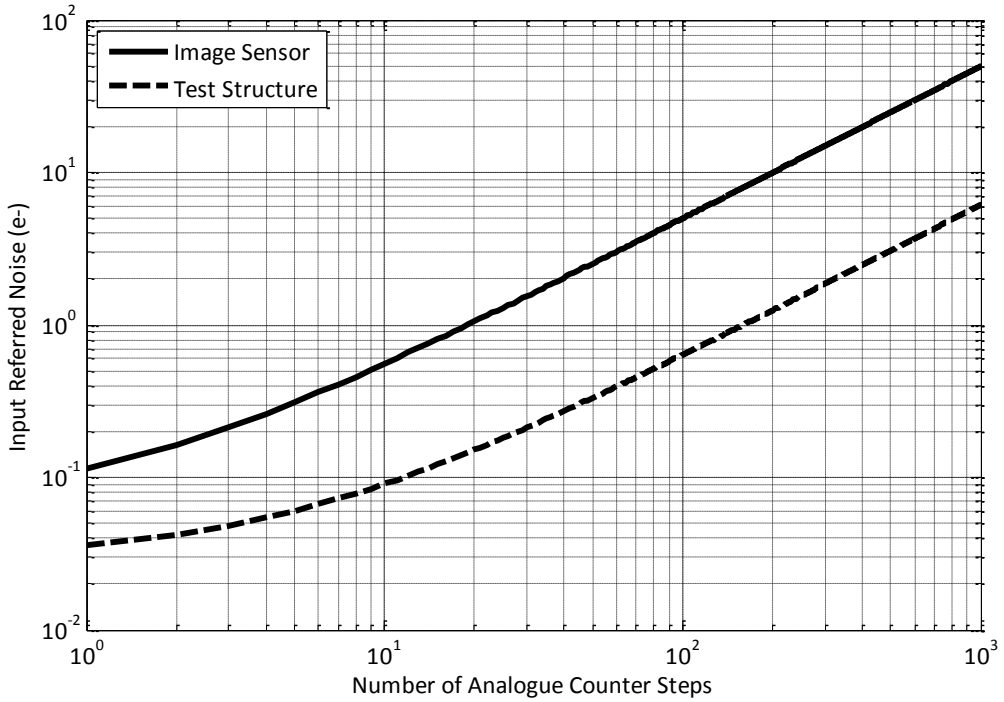


Figure 6.31. Extrapolated number of analogue counter events against the respective input referred noise.

Teranishi describes the condition for accurate or true electron counting at a maximum of 0.3e- read noise [164]. Fossum alternatively promotes that photon counting can be achieved with maximum 0.15e- read noise [61]. For the image sensor it can be calculated that the input-referred noise of the image sensor against this 14.2mV sensitivity is 0.06e- for the first SPAD event. Table 6.11 provides a summary of the number of analogue counter steps against the equivalent input referred noise.

Equivalent Input Referred Noise	Max. No. of Analogue Counter Steps	
	Image Sensor	Test Structure
0.1e-	1	11
0.15e-	2	19
0.3e-	5	45
1e-	19	160

Table 6.11. Equivalent noise at a range of cumulative SPAD events.

6.3.7. Full Well

The maximum counting capability or equivalent 'full well' is evaluated in an experiment by placing the sensor under uniform constant illumination from an LED. The exposure time was incrementally increased and the exposure was captured for all pixels. This experiment was performed for four settings of VS bias (200 to 500mV). Traditionally the saturated full well of a CIS or CCD pixel is obtained experimentally by the measurement of the maximum SNR [46]. However, in cases where the readout causes non-linearity in the upper portion of the output swing, the linear full well is calculated at the point where the output integral non-linearity (INL) deviates to 3% (i.e. the point at which the output voltage departs from a linear fit line by 3%) [165]. The latter metric of linear full well is used in this experiment. Figure 6.32 details the average pixel output voltage for the four VS bias settings against exposure time. The red marker dot on each graph (a) to (d), indicates the calculated linear full well from the measured INL. These data are tabulated in Table 6.12, alongside the fitted sensitivity values from Figure 6.21 against VS bias. The final column in the table calculates the equivalent linear full well as an integer number of SPAD events.

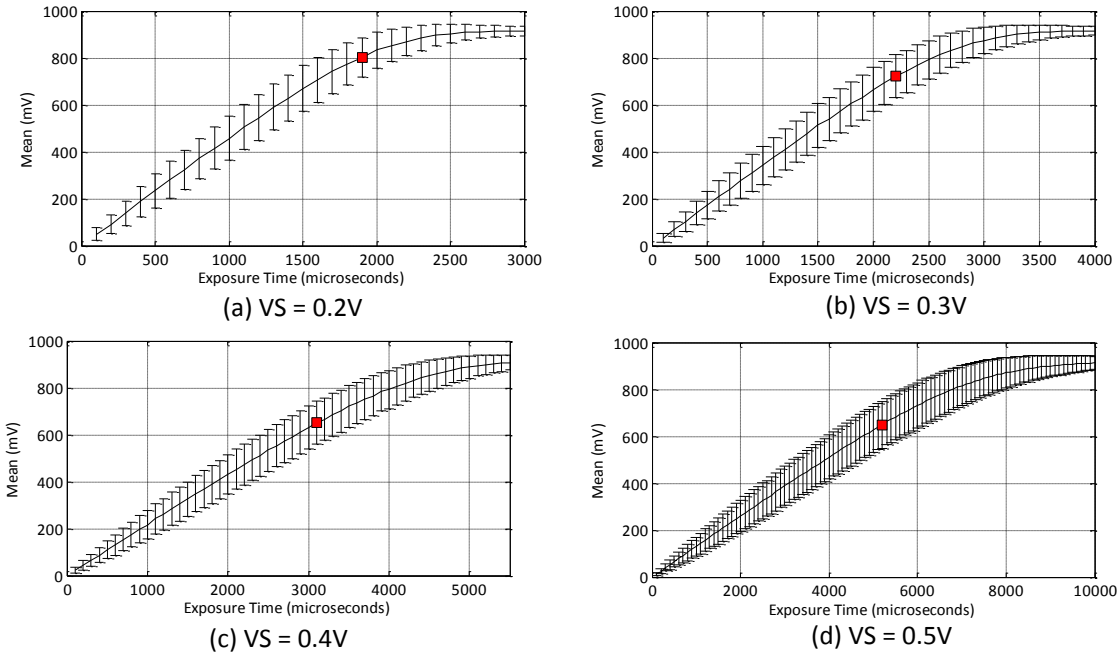


Figure 6.32. Full well experiment with increasing VS bias from (a) to (d). The red dot in each graph indicates the linear full well.

VS Bias (mV)	Linear Full Well Voltage (mV)	Fitted Sensitivity (mV / SPAD Event)	Equivalent Linear Full Well (SPAD Events)
200	802.8	14.26	56
300	722.1	11.23	64
400	651.4	8.21	79
500	648.3	5.19	125

Table 6.12. Equivalent linear full well table with increasing VS bias.

6.3.8. PRNU

In Chapter 4 the PRNU (predominantly pixel to pixel CTA gain mismatch) is measured using the FPGA pulse test method. In the image sensor this facility is removed. To measure the image sensor PRNU, a high number of SPAD pulses are required to observe the voltage deviation (gain mismatch) of each of the pixels. To induce an observable mismatch would require exactly 100 SPAD events at 10mV steps to achieve the 10 to 20mV difference. As previously seen, after 10 SPAD events the cumulative noise dominates and the PRNU cannot be determined. Two comments are made, firstly PRNU is less significant than the cumulative noise and secondly not creating a pixel test column or row with an FPGA test input was a design oversight.

6.3.9. Signal to Noise Ratio

Unlike conventional CIS and CCD SNR graphs, there is no read noise limited region of operation in a true photon counting sensor. Using the modelled parameters developed for the variability of the CTA structure developed in Chapter 4, and adding a modelled 700μV/SPAD event cumulative noise in quadrature, the SNR of the modelled CTA can be graphed alongside an ideal noise-less photon counting system. Both PRNU and cumulative noise per analogue counting step, flattens the SNR response as seen in the black line Figure 6.33(a).

The SNR from each of the previous full well experiments is calculated and plotted in Figure 6.33 (b) to (e). The calculated linear full well is marked again with the red marker. Furthermore, converse to a traditional SNR graph, the photon shot noise limited section of the SNR graph is evident for the first SPAD events, as seen in the left hand side of each of the graphs (b) to (e). In Figure 6.33 (b) the increase of the measured data above the ideal photon shot noise limited SNR line is due to the non-linear output where the signal is being increasingly clipped. The decreasing step size from (b) to (e) indicates that the cumulative noise has a more significant effect which is self-evident as the ratio of cumulative noise to the step size increases for decreasing step size. It is clear there is no read-noise region of operation in these SNR plots. This data aligns with the previous noise modelling in both Chapter 4, and this chapter in Figure 6.31 and in Table 6.11.

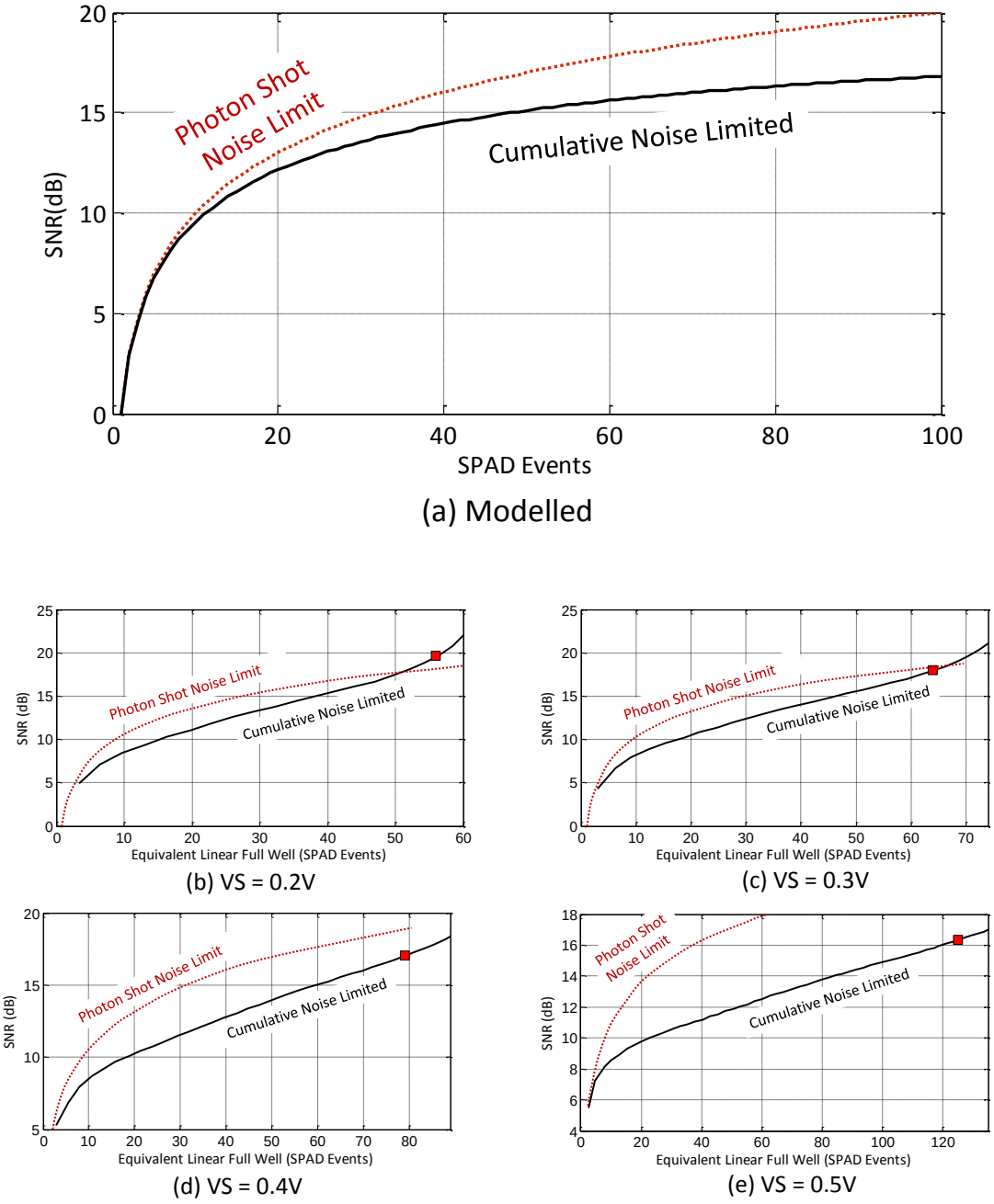


Figure 6.33. Signal to noise ratio (SNR). Modelled in (a) for ideal photon shot noise and with modelled cumulative noise. SNR for four VS biases (b) to (e) against the ideal photon shot noise limit. The calculated linear full well is marked with a red marker.

6.4. Digitally Oversampled Binary Image Sensor

The imager is demonstrated in this section, operating as a global shutter oversampled binary image sensor or SPAD-based QIS, reading out at kilo-FPS. Binary bit planes or field images are accumulated in real time on FPGA to construct multi-bit output frame images.

6.4.1. Quanta Image Sensor DLogH

Hurter and Driffield in 1890 conducted various experiments to quantify the photo to chemical response of photographic silver halide film [60]. The density of the exposed film (a measure of film darkening) was recorded under increasing exposure. Their famous density 'D' to normalised exposure 'E' is a logarithmic S-shaped curve, and a reproduction from 1923 is shown in Figure 6.34 [166]. Their experiments proved that film has a non-linear response, where unlike conventional CCDs and CISs saturating at high exposure, film has a natural compression of high-lights. Fossum in [61] showed that the fundamental operation of the QIS produces a similar logarithmic response with an S-shaped curve.

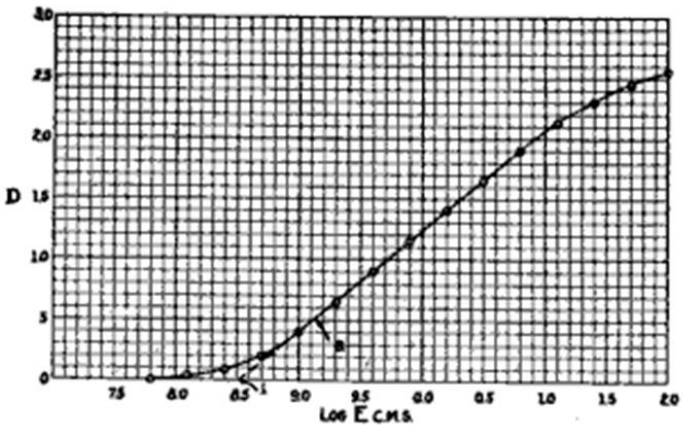


Figure 6.34. Density to exposure D-Log H curve of photographic film, from [166] **in 1923.**

The density in a QIS is the number of jots reporting a photon (or more than one) with logic '1' versus no photon and logic '0'. The exposure time (τ) is normalised by the average arrival rate of photons (φ) into normalised number of photons captured or 'quanta exposure' (H) where $H = \varphi . \tau$. The ideal QIS response, with a single photo-electron triggered jot (as opposed to multiple photo-electronics triggering the jot), is derived in [61] as a probability (P) of receiving one or more photons arrivals (k) within an exposure time (H):

$$P(k > 1) = 1 - e^{-H}$$

(Eq. 6.6.)

The QIS performance is measured by placing the sensor using constant illumination, incrementing the field image exposure time. The FPGA recorded an 11-bit frame image with 2048 field images temporally oversampled. The bit plane density 'D' (number of jots registering a SPAD event) is normalised, and the exposure time is normalised to quanta exposure 'H' using φ= 0.27, equating 3.7μs field exposure to 1.0H. Figure 6.35 illustrates the measured bit density versus exposure 'H', the ideal 'D-Log H' S-shaped curve QIS response from Equation 6.6, and the ideal linear response. The measured data have a good fit to the ideal QIS response. At full exposure (H=1.0) the linear line has saturated, and both the ideal QIS response and sensor response reach only ~63% density. This shows the compression of high exposures and of high-lights that this sensor achieves. The overexposure latitude is defined in [61] as the ratio of 99% of the linear exposure maximum to the QIS bit plane density reaching 99% of maximum. The measured data have a 4.6x overexposure latitude which matches QIS theory.

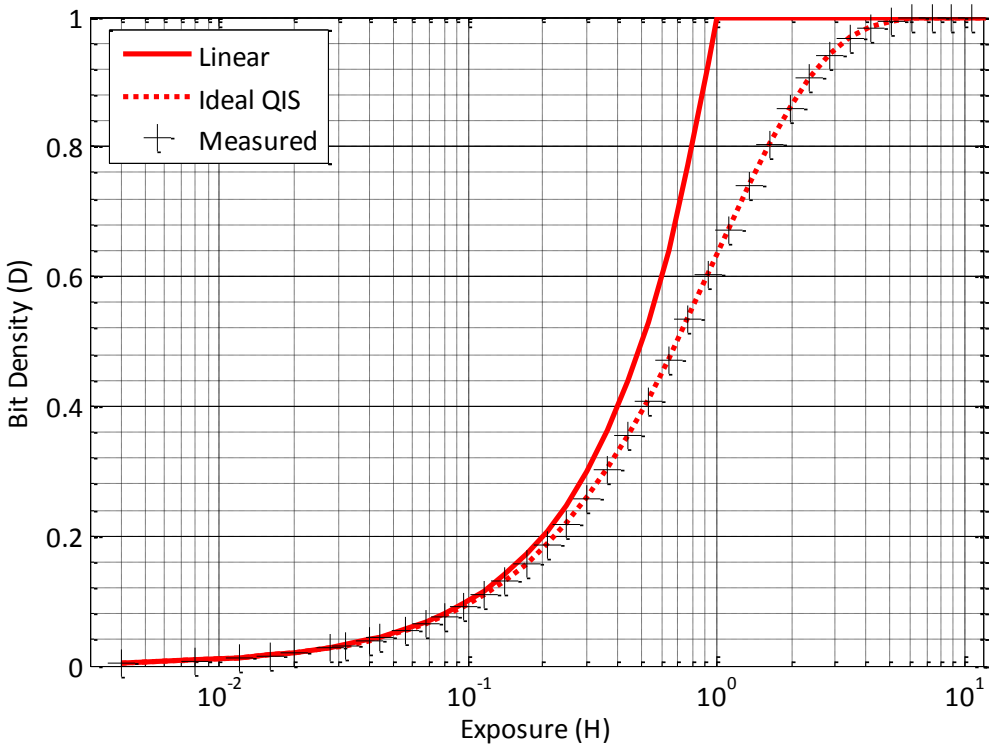


Figure 6.35. Ideal linear response and the D-Log H measured and ideal response of normalised QIS bit plane density versus normalised exposure.

6.4.2. QIS Noise

Figure 6.36 shows the measured signal and one standard deviation noise versus normalised exposure on a log-log graph. This highlights that at lower exposures the sensor is photon shot noise limited and at higher exposures (H>1.0) the shot noise is compressed as expected in a QIS. The deviation from the ideal noise is calculated and normalised against the mean bit density and an average is computed which represents the bit error rate (BER). The average BER is 1.7 x 10^{-3}, and using Fossum's BER to read noise conversion from [61], replicated in Equation 4.24, it produces an effective read noise of 0.168e-.

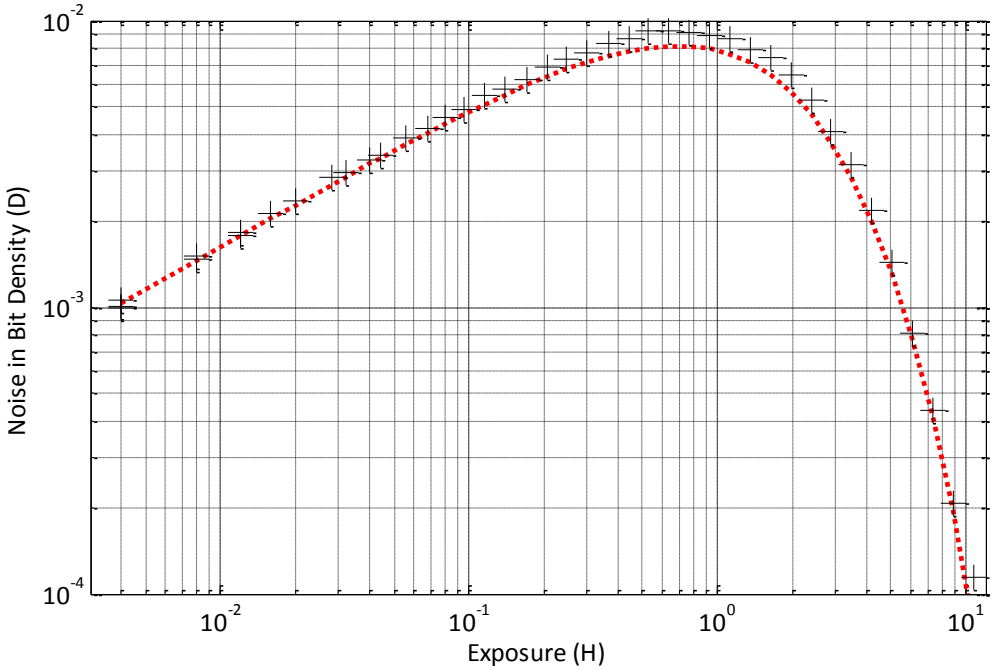


Figure 6.36. One standard deviation of recorded bit density versus normalised exposure.

The signal to noise ratio is plotted in Figure 6.37, showing 18dB SNR at H=1.0 and increases to 54dB at H=10.0. The SNR plot demonstrates only photon shot-noise limited response at lower exposures (H<1.0) with no read noise limited region of operation. With H>1.0, the shot noise is compressed in the over exposure region.

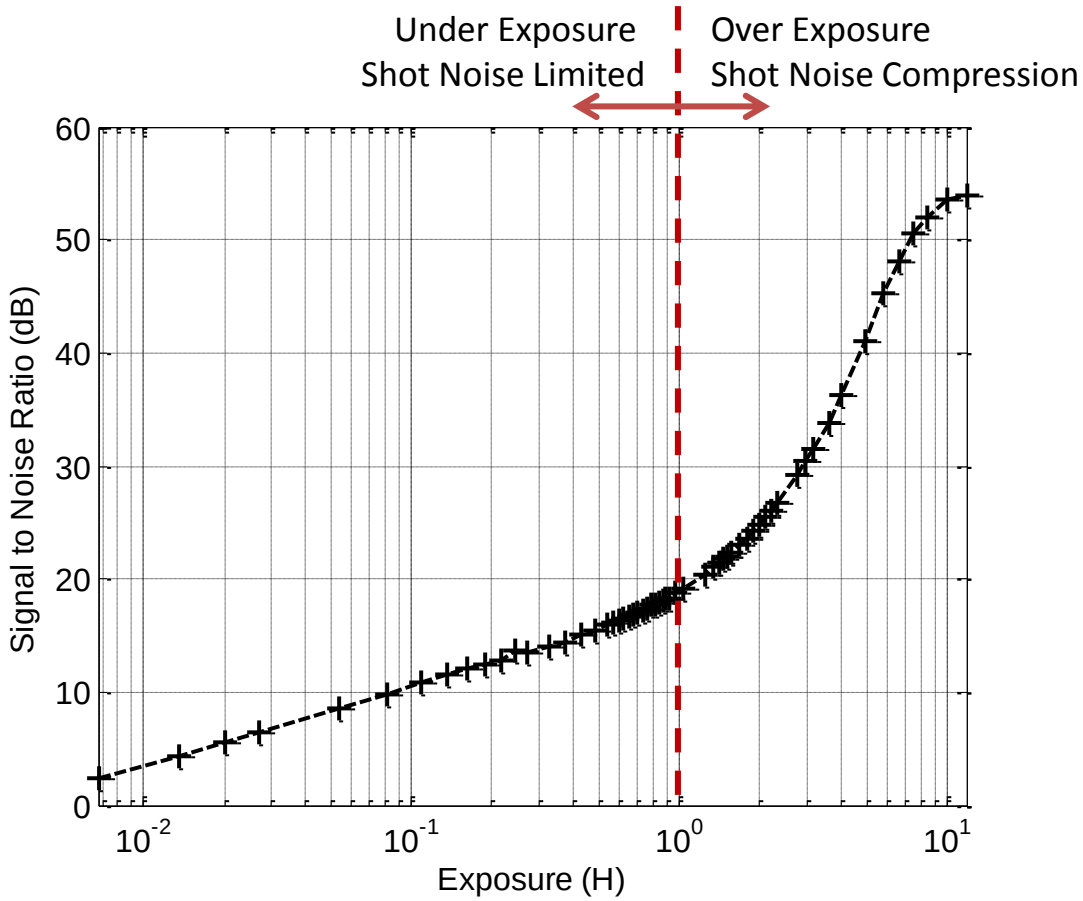


Figure 6.37. Signal to noise ratio of the SPAD-based QIS

6.4.3. Digital Oversampling Image Formation

A binary field image captured from the image sensor is shown in Figure 6.38 with a 1μs exposure. Figure 6.39 shows six temporally oversampled frames, incrementally doubling the bit depth of the output frame and halving the output frame rate.

Figure 6.40(a) illustrates an uncorrected 12b output frame with evident high noise pixels and (b) shows the image passed through a conventional ISP threshold median filter algorithm to remove bright pixels. High frequency spatial aliasing is evident in the background of these images as an optical low pass filter was not used.

Figure 6.38. Binary field image, 1μs exposure, F#2.0 lens, captured at 5,120 FPS.

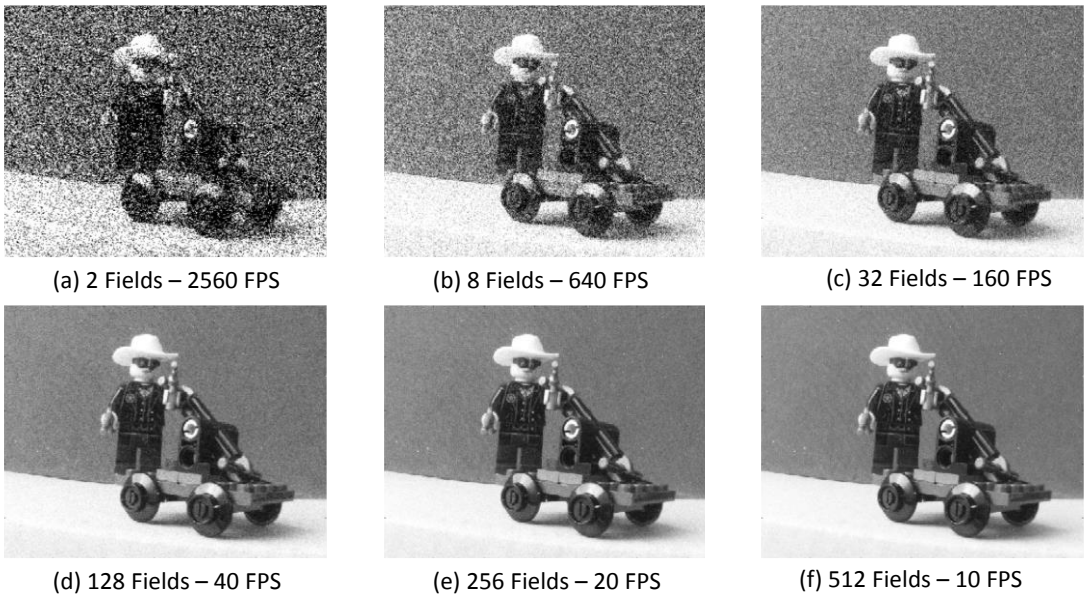


Figure 6.39. Digital oversampling in the time domain, accumulating successive field images to create an output frame. Six frame bit depth (2b to 9b) (a) – (f) are shown alongside the respective output frame rate.

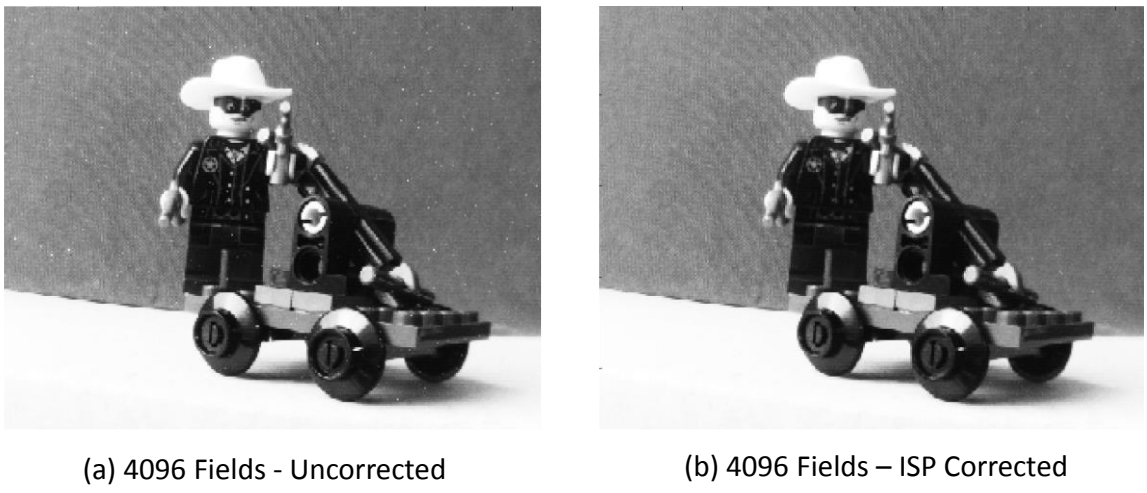


Figure 6.40 (a) A 12b output frame with evident high DCR SPADs. (b) Conventional ISP threshold median filter applied to remove high noise pixels. The contrast in both images has been increased to highlight the high DCR pixels.

6.4.4. FPN in Oversampled Images

The column wise readout performs a coarse flash conversion. The pixel noise mechanisms and offsets (kT/C noise, flicker noise, source follower Vt variation, etc.) once input referred, are rendered insignificant by the high conversion gain of the SPAD and the memory discharge. Hence in single bit SCS mode, the need for CDS is removed and the data conversion is performed in a single step. However, the global voltage reference to the differential dynamic latched comparator must be set to ensure correct conversion between reset level and the memory discharge level.

Figure 6.41(a) and (b) demonstrate the output field and frame images for the voltage reference set too high (~1.6V) around the pixel reset level. VFPN, due to pixel source follower Vt variation, is dominant in both field and oversampled frame image. Figure 6.41(c) indicates the correct voltage reference is set (~1.0V). The frame image in Figure 6.41(d) is median filtered to remove high DCR pixels and no FPN is evident or observable.

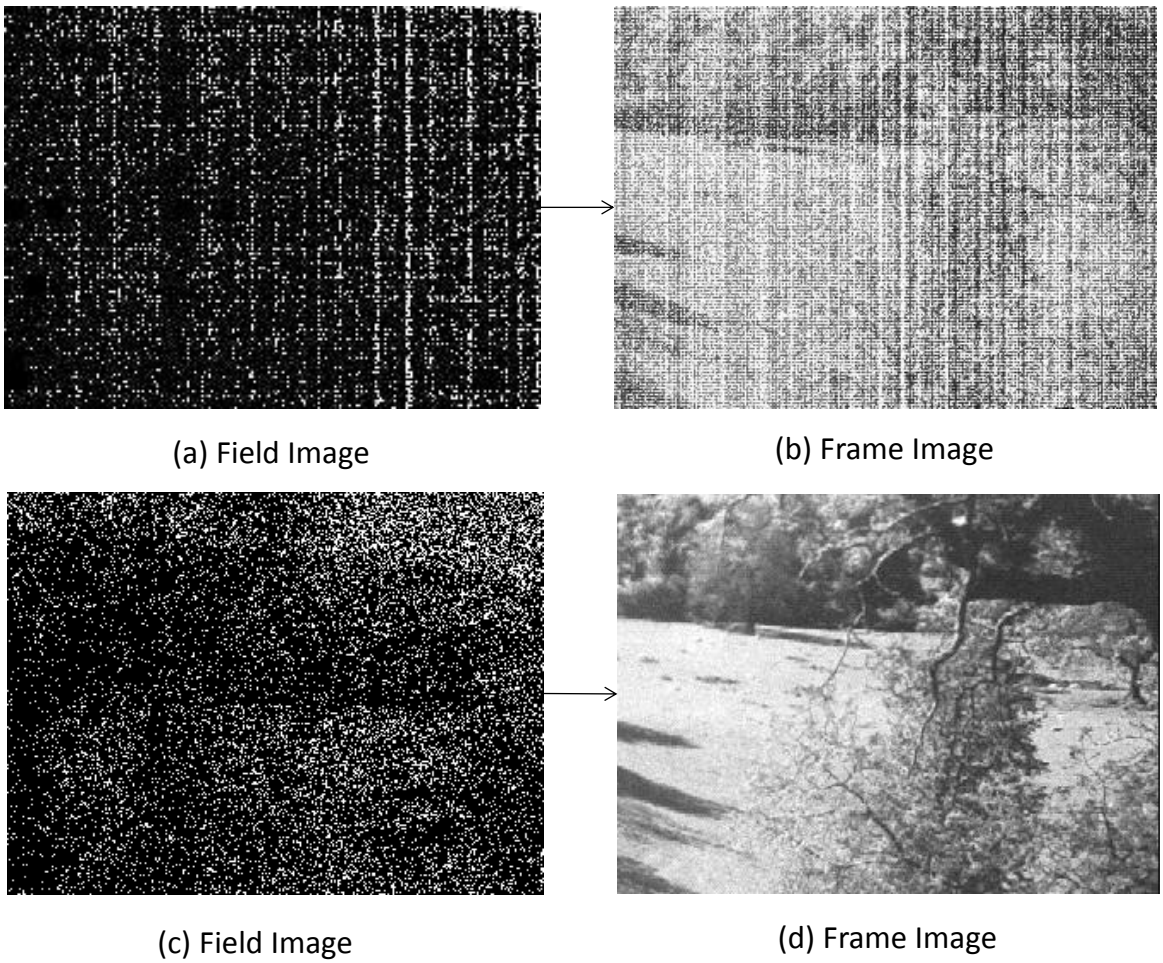


Figure 6.41 (a)-(b) VFPN dominant in field and frame images due to incorrect comparator voltage reference setting, (c)-(d) Correct voltage reference setting with no determinable FPN.

6.4.5. Digitally Oversampled TOF Frame Rate

In gated and oversampled ITOF in QIS mode, the sensor captures single bit field images with a number of laser repetitions per field. Table 6.13 shows the achievable output oversampled frame rate for 1-bin gated and 4-bin oversampled ITOF imaging, with an example 160 laser repetitions per field comprising a 35.5µs field exposure. The field rate of the sensor with 45MHz clock rate, for time-gated or digitally oversampled TOF imaging is 9072 fields per second. To achieve >30 FPS output frame rate of gated images, it would require oversampling less than 300 field images, with 8b frame depth or less. However, TOF imaging with ambient subtraction requires 4 gated field images per frame image decreasing the output frame rate by a factor of four. Rolling shutter timing has not been implemented in this work, although this is shown theoretically in the table showing the slight increase in frame rate achievable.

Clock Frequency	45 MHz			
Clock Period	222ns			
Laser Repetitions Per Field	160			
Field Readout Clock Cycles	4800			
Field Rate (FPS)	908			
Oversampled Fields per Frame	**256**	**512**	**1024**	**2048**
Oversampled Frame Bit Depth	8	9	10	11
Clock Cycles per GS Frame	1269760	2,539,520	5,079,040	10,158,080
Clock Cycles per RS* Frame	1228800	2,457,600	4,915,200	9,830,400
GS Gated Image – 1bin Frame Rate (FPS)	**35.44**	**17.72**	**8.86**	**4.43**
RS* Gated Image – 1bin Frame Rate (FPS)	36.62	18.31	9.16	4.58
GS ITOF Image – 4bin Frame Rate (FPS)	**8.58**	**4.29**	**2.15**	**1.07**
RS ITOF Image – 4bin Frame Rate (FPS)**	9.16	4.58	2.29	1.14

Table 6.13. Frame rate calculations of the SPCIMAGER 'AB' capturing digitally oversampled 1-bin gated and 4-bin ITOF images. *Theoretical rolling shutter timings for comparison.

6.5. Time Gating

6.5.1. Pulse Generator Characterisation

The two pulse generators included on-chip create the two time-gate pulses ('A' and 'B'). The pulse generators consist of a tapped delay line, where each delay cell is current starved, controlled by a global non-linear 4-bit DAC. These are electrically characterised for the mean temporal width of a single delay element (or tap) in the delay line, and the total dynamic range (or maximum achievable delay) over the range of sixteen DAC settings. The sensor has two digital pads (signals 'BINAOUT' and 'BINBOUT' listed in Appendix 3) that are connected to the 'A' and 'B' time gate inputs of test row 255. This facilitates measuring the time gate pulses after progressing through the pulse generators and metal clock trees. The pulse generators from one sensor are measured using a LeCroy 'WaveMaster' oscilloscope triggered by the input FPGA signal and recording the electrical pulse width of the test output. Table 6.14 displays the calculated mean and standard deviations of the bin width, alongside the mean total dynamic range. DAC setting 15 is not allowed as it switches the current DAC off, to prevent this there is a protection function in the firmware which sets DAC setting 14 if the value of 15 is requested.

DAC settings 0 to 7 allow for approximately linear fine tuning of the pulse edge in 500ps increments over an approximate 60ns dynamic range. Whereas a non-linear DAC response was designed in settings 8 to 14 to cover a wide range, the pulse generators are measured attaining a dynamic range of 80ns to 473ns with 0.6ns to 3.7ns steps respectively. The large standard deviation in bin width of DAC setting 14 is due to the generated voltage being set at the threshold voltage of the current starving NMOS transistors, and through device mismatch, causing some to be sub-threshold and those delay elements to have much longer delay times than others.

DAC Setting	Mean Temporal Bin Width (ps)	Std. Dev. Temporal Bin Width (ps)	Mean Dynamic Range (ns)
0	482.6	30.7	61.8
1	483.7	38.6	61.9
2	487.9	33.4	62.4
3	491.3	33.2	62.9
4	500.5	42.6	64.1
5	505.9	34.7	64.8
6	514.0	36.8	65.8
7	520.0	42.0	66.6
8	626.6	47.5	80.2
9	656.2	46.9	84.0
10	720.9	65.3	92.3
11	786.1	56.8	100.6
12	1240.2	141.9	158.7
13	1619.8	202.5	207.3
14	3447.8	1396.3	441.3
15	3447.8	1396.3	441.3

Table 6.14. On-chip pulse generator characterisation results.

Instability in the pulse generator output is observed under certain short pulse and DAC settings, whereby the output flips polarity. This is attributed to the feed-back loop in the bang-bang phase detector failing to clear the two flip-flops. Subsequent pulses then create a short 'off' period, rather than the desired 'on' period. The flipped polarity of the pulse generator can be reset by selecting both delay tap outputs to the zero position and triggering it once. Future revisions of this sensor, or other sensors using this pulse generator block should review and fix this instability.

6.5.2. Time Gate Measurement

Figure 6.42 displays a diagram of the experimental setup used to characterise the time gating of the image sensor. The laser trigger output from the FPGA is connected to a picosecond delay generator (Stanford Research Systems (SRS) DG645) which in turn triggers a PicoQuant PDL800 pulsed laser 80ps FWHM, with 425nm laser head. To avoid SPAD paralysis due to the high intensity of the laser, the laser head is mounted beside the image sensor and is directed towards a white card through a diffractive diffuser. The sensor is uncovered with no lens ensuring all pixels are uniformly illuminated. Both sensor and delay generator are computer controlled. The delay of the pulsed laser in relation to the time gate is incremented in 5ps steps. At each step an intensity image is captured and stored, in order to characterise the profile of the time gating pulse.

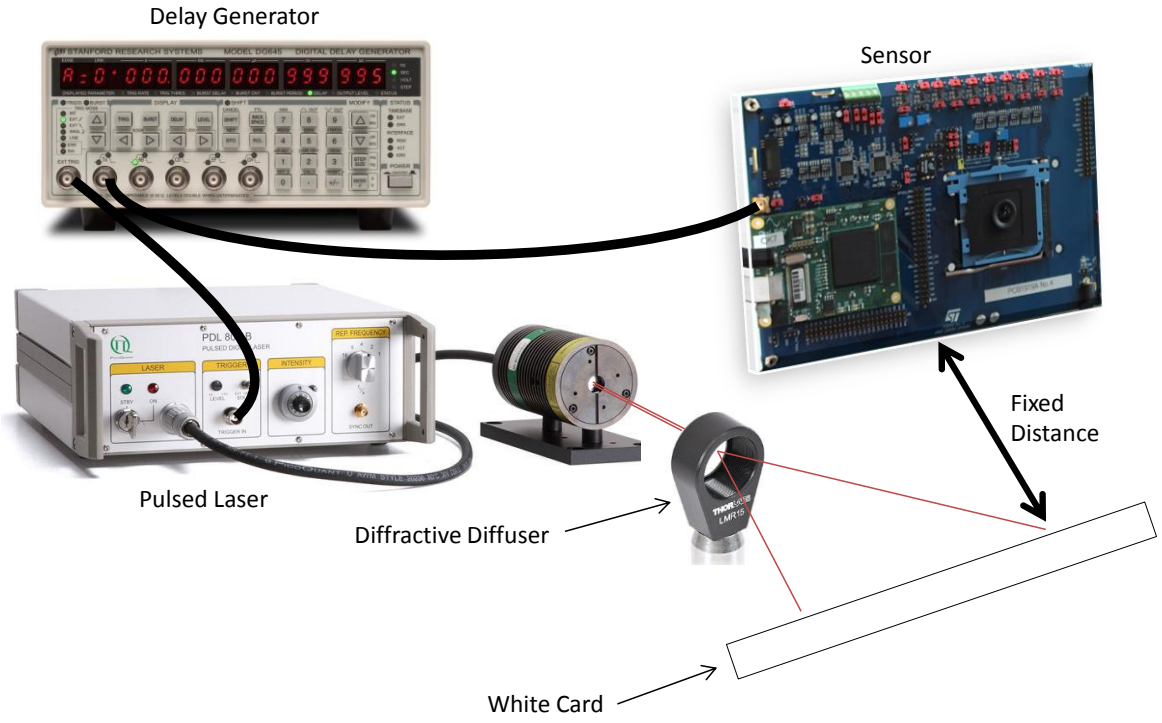
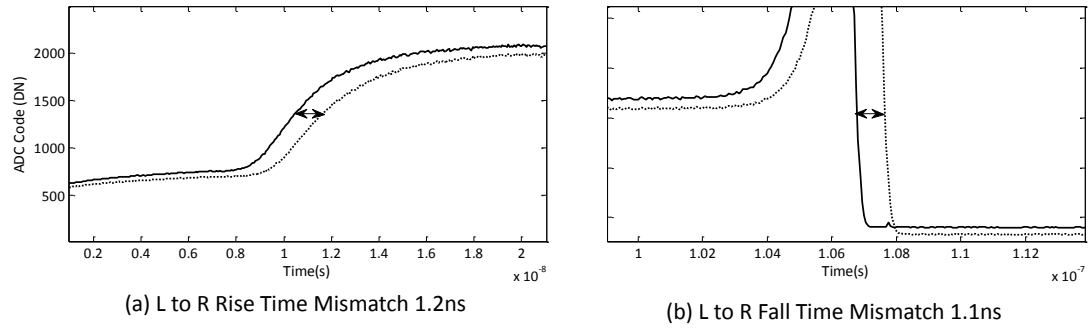


Figure 6.42. Diagram of time gating experimental setup.

6.5.3. SPCIMAGER_AA Time Gate Mismatch

The 'AA' sensor is set in analogue counting mode to record the intensity of the sensor at each delay increment. For each delay, ten images are recorded and averaged in post-processing. In the first IC revision 'SPCIMAGER AA', a left-half to right-half time gate mismatch is evident in the time gate measurements. Figure 6.43 shows this effect plotting the averaged output of the left-half to the right-half of the sensor displaying a 1.2ns mismatch in rise time and 1.1ns in fall time. This is equal to a range mismatch in ITOF images of 18cm which is undesirable. This discrepancy is corrected in 'SPCIMAGER AB' by removing the right time gate drivers. The remainder of the time gate and TOF measurements are performed with the 'AB' IC revision.



(a) L to R Rise Time Mismatch 1.2ns

(b) L to R Fall Time Mismatch 1.1ns

Figure 6.43. 'SPCIMAGER AA' left half (solid line) to right half (dotted line) time gate mismatch. 1.2ns mismatch in rise time, 1.1ns mismatch in fall time. This misalignment is corrected in the 'AB' IC revision by removal of the right half time gate drivers.

6.5.4. Time Gate Shape – Analogue Counting

The 'AB' image sensor is configured in analogue counting mode to record the intensity of the sensor at each delay increment. For each delay step, ten images are recorded and averaged in post-processing. Figure 6.44 illustrates the average sensor output showing the time gate shape of a 100ns gate duration. Four features are marked in the figure, however, the shape of the time gate is not the expected rectangular or trapezoidal, and evidently the 'spike' component does not fit in this ideal shape and requires explanation. In initial simulations and device verification, this effect was not seen. However, after investigation the 'spike' effect can be realised in simulation of the pixel time gate and counter, by delaying the onset of the time-gate disable after the falling edge of enable which is described in the following paragraphs.

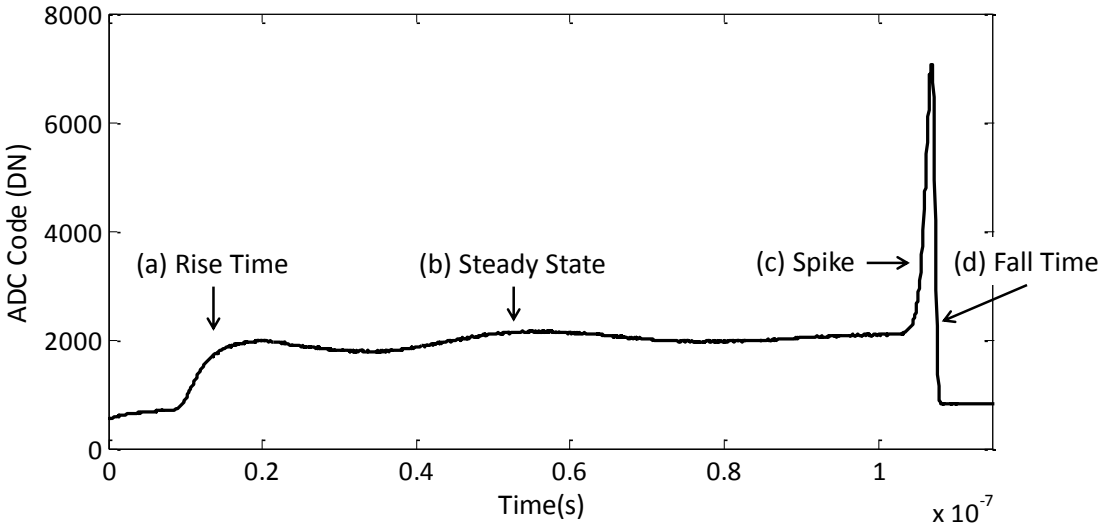


Figure 6.44. Time-gate shape of a single row captured using the pixel in analogue counting mode. Four features marked (a) to (d).

The time-gate front end (with disable M3 sub-threshold) acts as a sample and hold circuit, with the parasitic at the counter input as the capacitor and the time-gate enable M2 device as the switch. If a SPAD is mid-avalanche or recharging as the time-gate signal falls, then the parasitic will sample and hold the high SPAD anode voltage. The held voltage will remain until the disable signal is asserted. The CTA will trigger with the onset of the SPAD avalanche, yet the circuit is intended to function with the CTA 'VB' node discharge time greater than the SPAD dead time. This sample and hold effect reverses this situation: if the sample and hold effect occurs, the input to the CTA has a much longer discharge time than the 'VB' node. Effectively bringing the source-follower transistor M5 into weak inversion after the initial CTA channel cut-off operation, hence allowing charge to leak from the capacitor.

Ideally the disable rising edge should immediately follow the enable falling edge but be non-overlapping. Figure 6.45(a) illustrates this ideal minimised overlap at the falling edge of the time gate. On the image sensor, the disable row drivers are fed from a signal directly from the FPGA input signal with no timing balanced structures to the row-drivers, whereas the enable passes through the delay generator and a timing balanced clock tree. This was an oversight in the design as it causes a prolonged falling edge overlap as shown in Figure 6.45(b).

To evaluate this distortion, a simulation is performed incrementing the onset of the disable rising edge to the falling edge of the time gate and recording the CTA counter voltage. Figure 6.46 displays the simulation with a 30ns falling edge overlap and matching the experimental conditions of the time-gate measurement by incrementing the onset of a single SPAD pulse. The spike is clearly evident in simulation, at the end of the time gate period, confirming the experimental data.

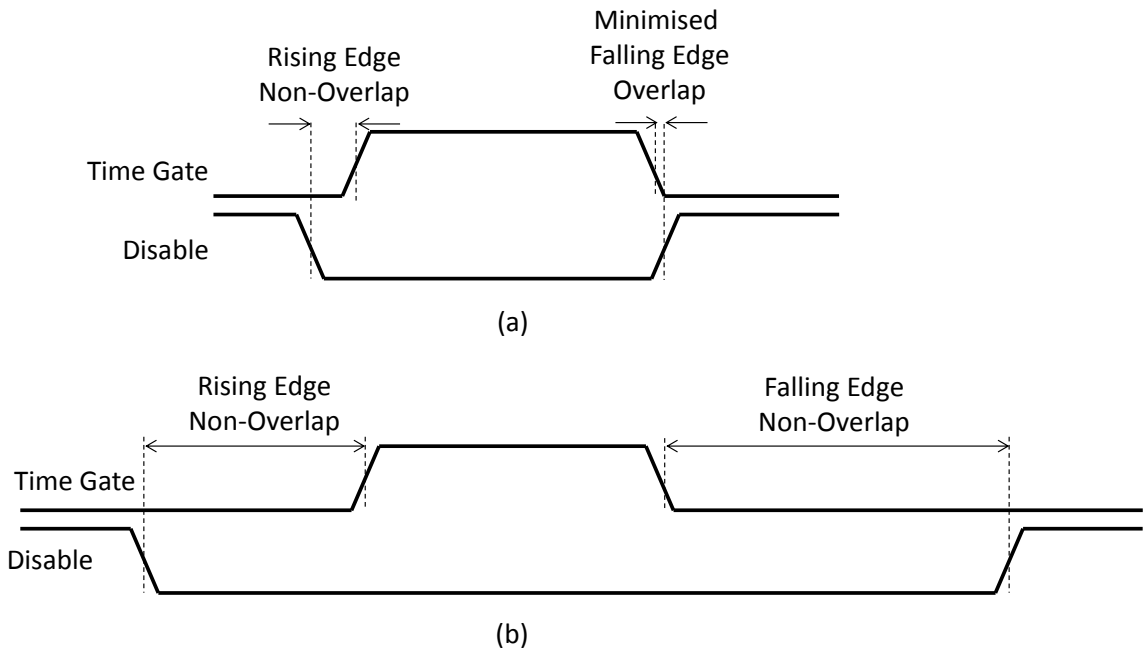


Figure 6.45 (a) Ideal time-gating with minimised overlap of time gate falling edge to disable rising edge (b) Prolonged time gate falling edge non-overlap as found on the image sensor.

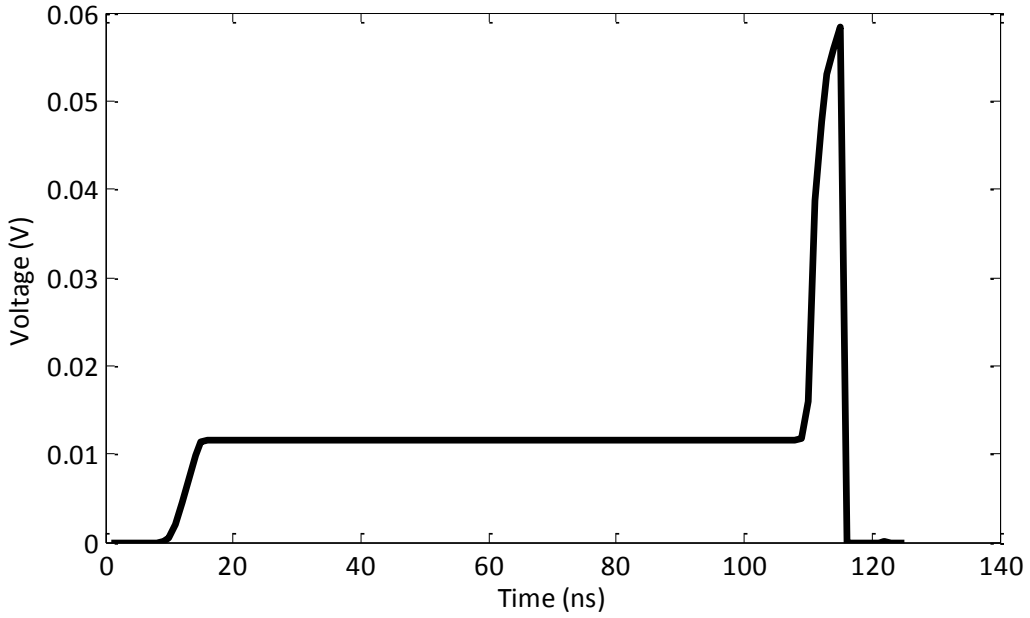


Figure 6.46. Simulated time gate shape of the pixel's analogue counting mode for a 30ns delay or non-overlapping period between the onsets of disable rising edge to the falling time gate edge.

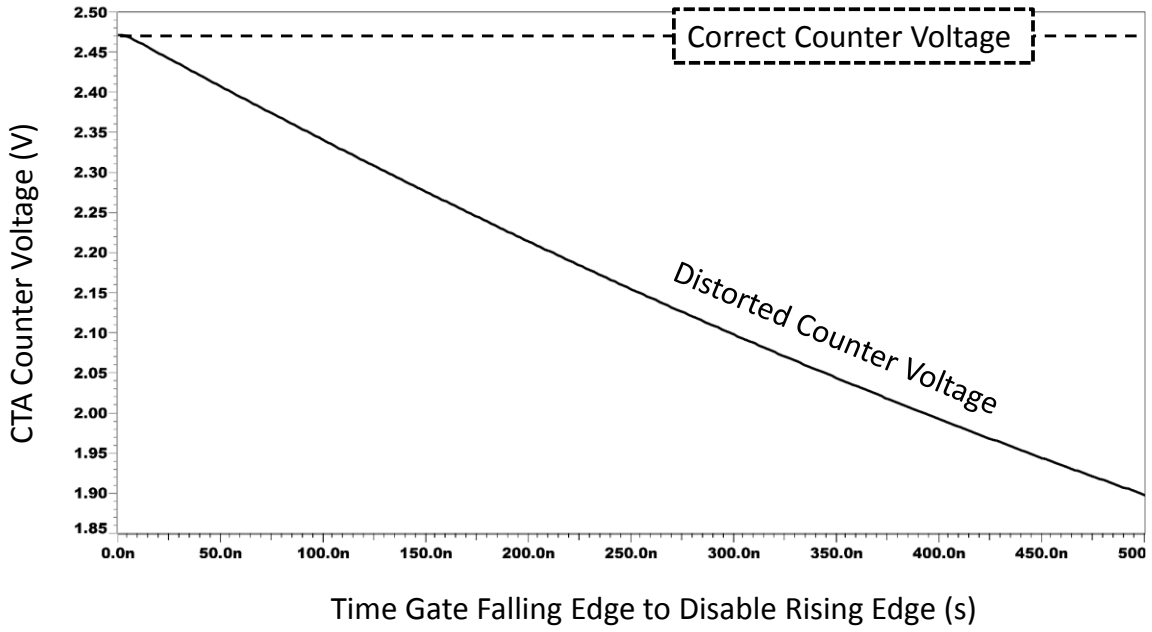


Figure 6.47. Simulated effect on the CTA voltage from an incremental non-overlap period between time gate and disable pulses.

A second simulation was performed aligning the onset of a SPAD pulse, with 10ns dead time, 1ns before the time gate falling edge. Figure 6.47 illustrates the relationship between the CTA counter voltage and the non-overlapping period between the time-gate falling edge and the disable rising edge. The correct counter voltage is maintained for a non-overlapping period of less than 2ns otherwise the counter voltage is distorted.

6.5.5. Time Gate Shape – Digital Oversampling

The same experiment is conducted to evaluate the shape of the time gate in the digital oversampling mode. The advantage of this mode is that the sample and hold effect assists the discharge of the memory when a SPAD event occurs close to the time gate edge. The digital mode is tolerant of the time-gate issue, where the analogue counting mode is not.

6.5.5.1. Row Uniformity

Figure 6.48 illustrates two averaged sensor row outputs captured from a sequence of digital TOF image captures showing the time gate shape of an approximate 30ns gate duration. There is an 8ns mismatch of the onset of the centre row (120) to the outer row (1). However, both the ripple in the steady state and the fall-time spike seen in the analogue mode are not present. The calculated mean rise time and fall times of the time gating are presented in Table 6.15.

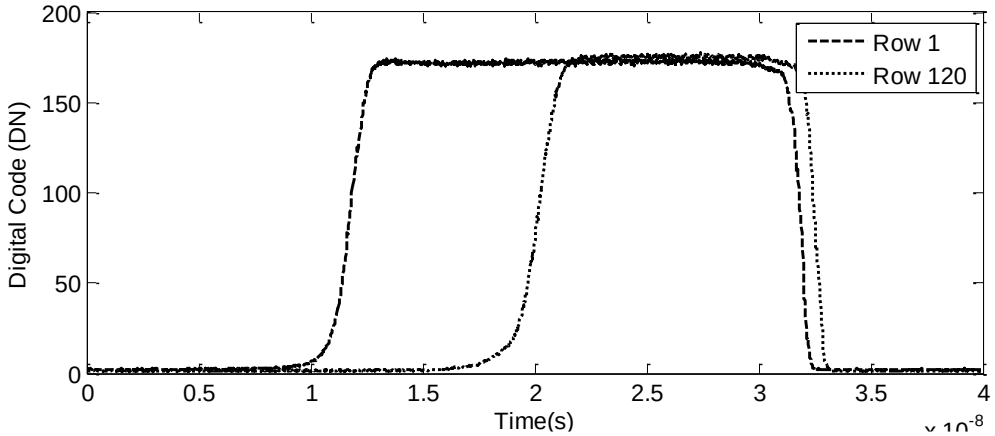


Figure 6.48. Time gate shape captured with Digital TOF image capture. Two row averages are shown, the outer row '1' and the centre row '120'. The rising edges have 12ns mismatch and the falling edges have approximately 750ps mismatch.

	Calculated Mean from 10% to 90% of Waveform	Standard Deviation in Rise and Fall Times
Rise Time	2.100 ns	191ps
Fall Time	0.997 ns	66 ps

Table 6.15. Calculated mean rise and fall times of the sensor time gates.

To evaluate the rise and fall time mismatches across the sensor, a row average is calculated per image frame at each delay step, and plotted versus time in Figure 6.49. The 8ns rise time mismatch is identifiable in the leading edge bow shape. This is attributed to a row driver local power supply collapse whilst the level shifters are switching high. The IC layout of the row driver and level shifter layout has thinner, non-gridded connections on the power supply single track and thicker, gridded connections on the ground supply. The falling edge also has a bow shape but not as significant.

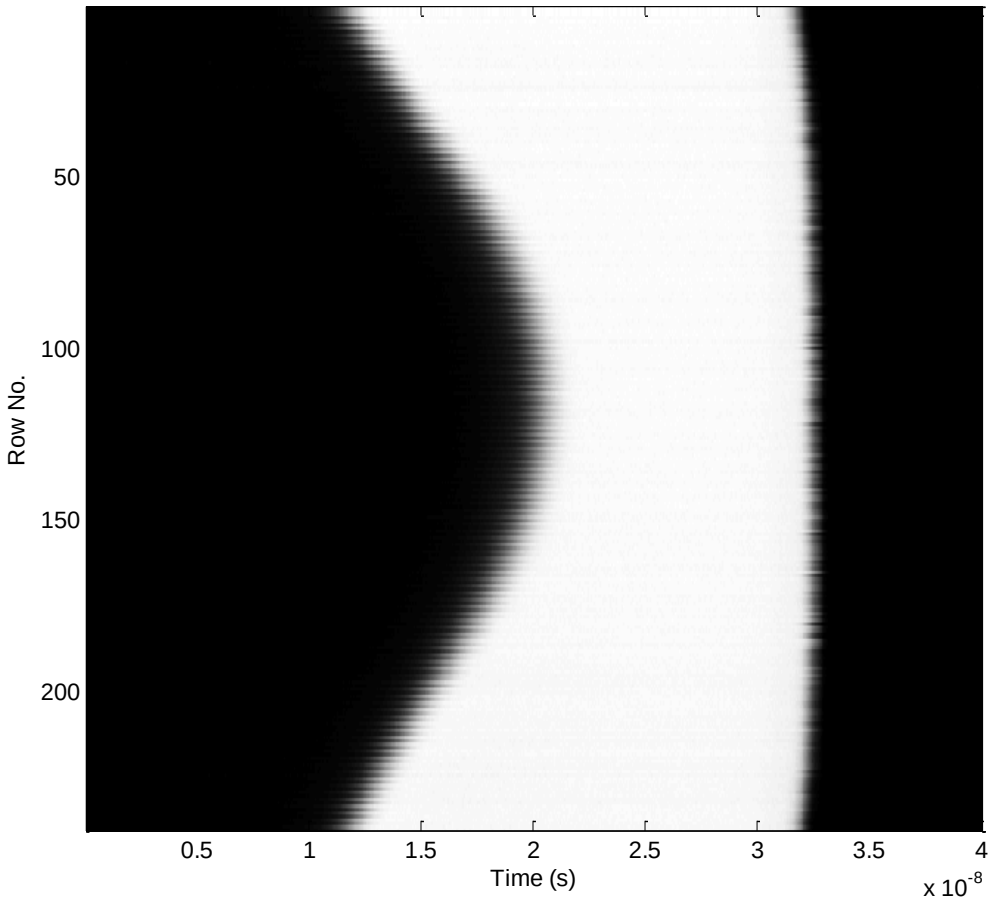


Figure 6.49. Time gate row driver mismatch. Each row line is an average of all pixels in that row.

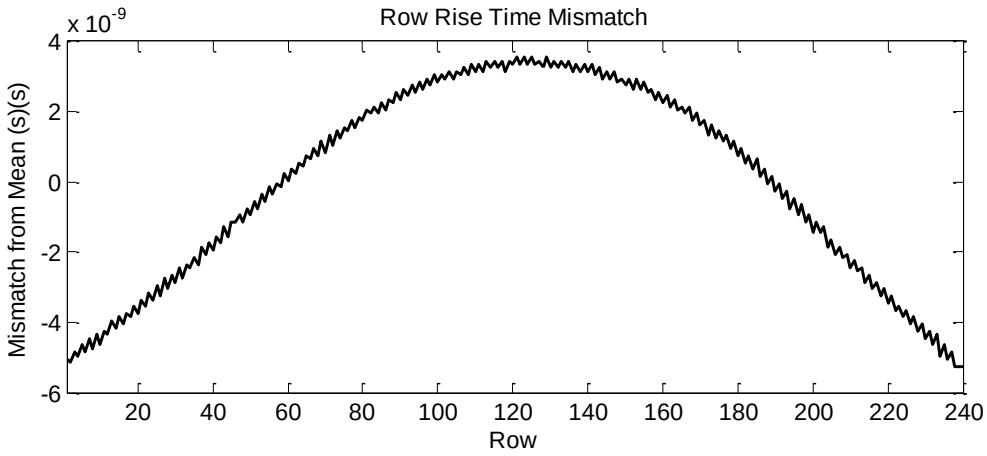


Figure 6.50. Rise time mismatch calculated at 50% crossing threshold of the average row signal.

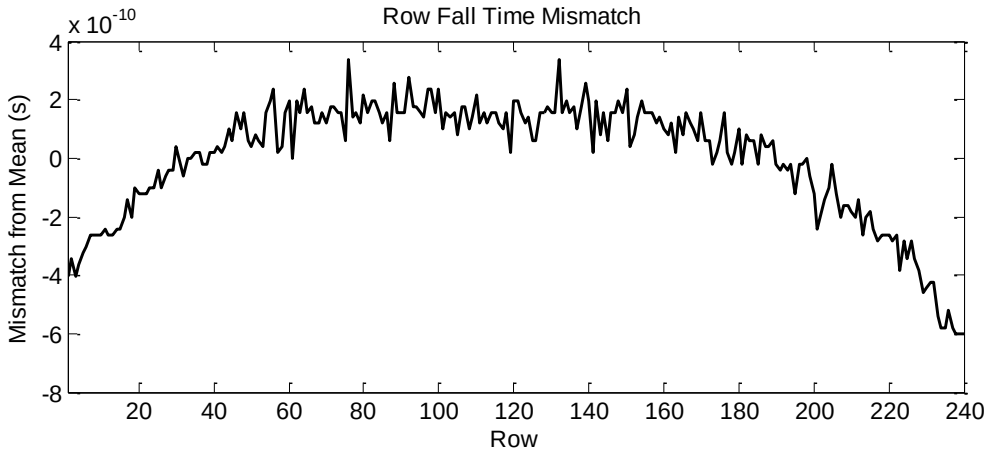


Figure 6.51. Fall time mismatch calculated at 50% crossing threshold of the average row signal.

The mismatch in the rising and falling edges of the time-gate are evaluated at a 50% threshold in signal of the averaged row signal. The rise and fall mismatches of the time-gate row drivers are plotted in Figure 6.50 and Figure 6.51 respectively.

Figure 6.51 has a comparable bow shape from the power supply collapse during the time gate pull-down although over a 900ps worst case rather than 8ns. On top of the bow shape, the row-to-row mismatch of the balanced clock tree routing is evident. The mismatch of row 1 approximately equals the mismatch of row 225 and not 240 (15 rows from the edge), as the 15 row drivers of the test pixels are located above row 0 and cannot be disabled. Row 1 is in reality the sixteenth row on the sensor, due to the fifteen test pixel rows.

6.5.5.2. Falling Edge Mismatch

Rise and fall time mismatch induce TOF range errors in computed TOF images. For ITOF image capture, the 8ns rise time mismatch is unusable. Therefore, ITOF image capture must be operated using the time gate falling edge configuration. The fall time mismatch across the whole array is calculated at the 50% crossing threshold. Figure 6.52 illustrates the temporal mismatch of the time gate falling edge, coloured from 0 to 1ns. Yellow and red areas of the image are delayed in respect to green and blue regions. The individual or discrete dots in the image represent those pixels with time-gate switches whose voltage thresholds are different from their surrounding neighbours.

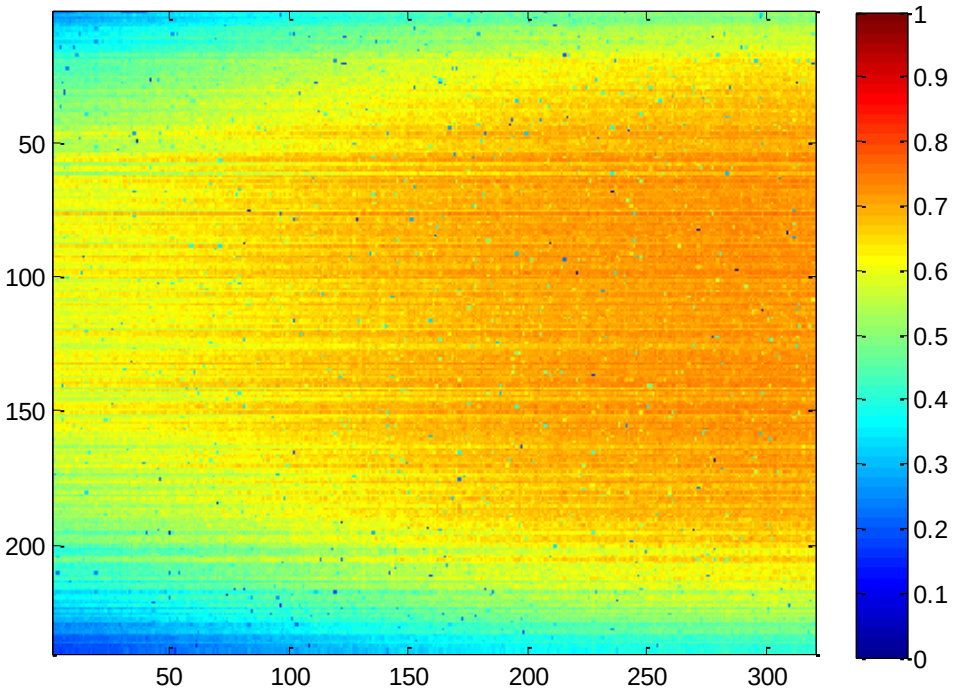


Figure 6.52. Temporal mismatch of the time gate falling edge. The image colour scale is in nanoseconds.

6.6. Time of Flight Imaging

This section highlights the TOF imaging in only single bit digital oversampling mode. Although the analogue counting mode is demonstrated for single photon counting, the image sensor is not successful at ranging with both issues of the time gate rising edge 'bow shape' and the falling edge spike precluding obtaining TOF imaging in the analogue mode.

6.6.1. Digitally Oversampled ITOF Ranging

Figure 6.53 illustrates the experimental setup for capturing TOF ranging with the optical breadboard, laser, laser diffuser optics and the image sensor highlighted. A small optical breadboard is mounted on a computer-controlled motorised rail to mount the image sensor, laser head and optics. To capture TOF range using the digital oversampling mode, two time-gated images, 'A' and 'B', are captured in an interleaved fashion. For ambient rejection, four fields are sequentially captured, the first two fields capture time-gated A and B images, and the second two images repeat the time-gating image capture with no laser. For ITOF capture with no ambient rejection the second two fields are not captured. The limited size of the RAM on the Spartan 3 FPGA forces the maximum width to be 18b, which is split into two 9b counters for accumulation. For two-bin acquisition with only up counting (with no ambient rejection) there is a maximum limit of 512 accumulations.

The ambient rejection mechanism with both up and down counters, extend this maximum 512 field acquisition limit and it is used throughout the remainder of these measurements. The readout is only 16b so the LSB of each 9b bin value is truncated.

The image sensor is secured to a fixed post. For range measurements, a diffractive optical diffuser with 20 degree dispersion is placed in front of the sensor as shown in the figure. It is removed for image capture. The PicoQuant 850nm laser head is placed co-incident with the image sensor on a flexible mounting allowing adjustment in all three axes. The PicoQuant nanosecond pulse width laser driver (FDL 500) is set to produce a 4ns pulse length optical stimulus. The laser driver is set in external trigger or slave mode triggered by the FPGA as timing master. The 2.4m length rail is powered by a National Instruments (NI) single axis motor controller controlled by NI 'LabView' software. The system can provide millimetre increments in separation of the target from the image sensor.

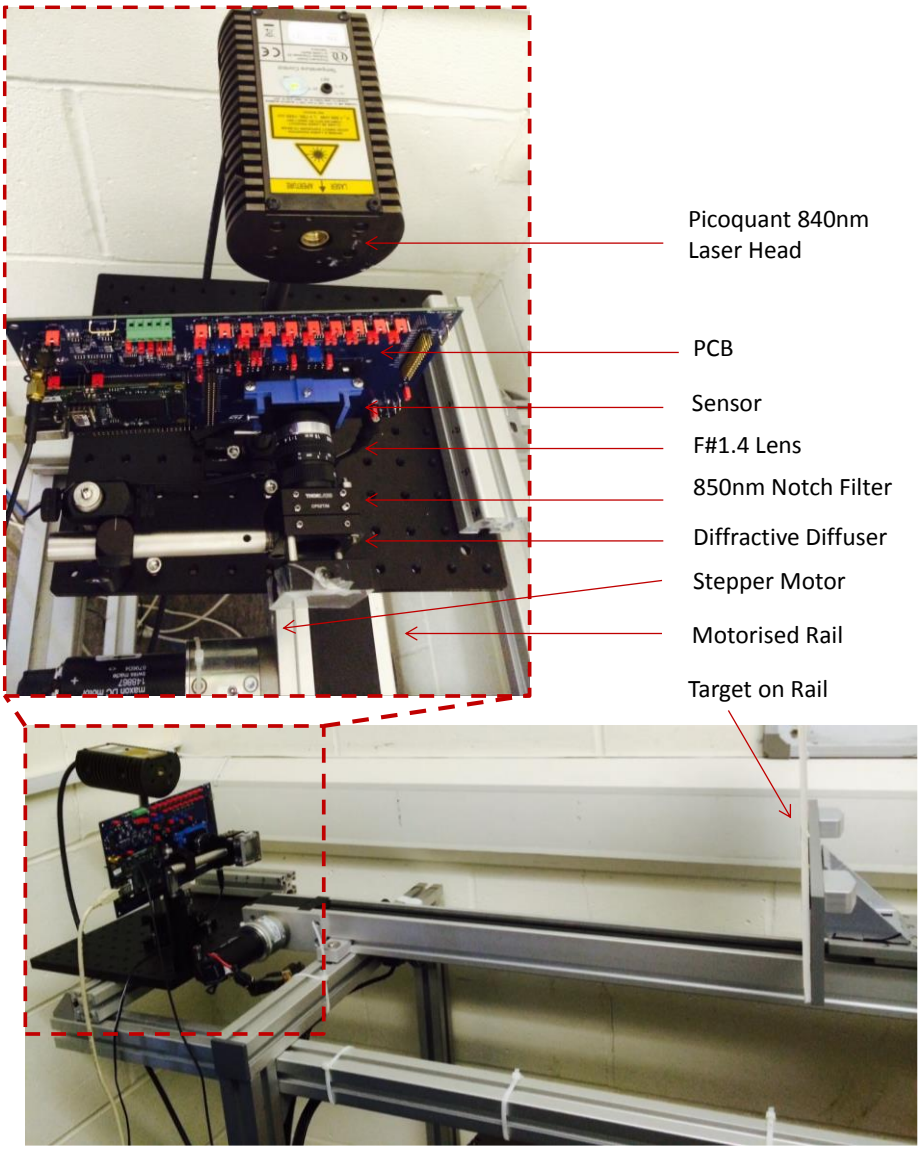


Figure 6.53. Ranging experimental setup mounted on a motorised linear rail.

The ITOF ranging performance is characterised with the sensor and a PicoQuant FDL-500 840nm 4ns pulse-length laser mounted on a linear motorised rail with 88% reflectance flat white target in dark room conditions. Figure 6.54 shows the mean reported range (Z) versus rail position using a 4ns laser pulse.

The temporal alignment of the time-gates with the laser output is a manually iterative process. The 'zero' mm distance is a few cm in front of the sensor as picosecond calibration is not possible. Targets around the 'zero' mm calibrated distance have high reflectance and saturate the sensor, and furthermore with the optical alignment of the laser head above the sensor it is difficult to evenly illuminate a target at this close range. The ambient rejection evidently functions to mitigate the effect of high DCR SPADs affecting the furthest target distances with low reflectance and low SNR. The measured distance with the ambient rejection disabled suffers from distortion due to the low differential, high common-mode, ITOF signal. Ambient noise could easily make this distance report above the ideal line as well as below.

The standard deviations and mean errors in Z are shown in Figure 6.55. There is an increase in range error at low distance which is attributed to the QIS overexposure latitude which at high light level is compressing the received TOF signal and distorting the calculated range. No correction is made for the logarithmic QIS response. The precision, at a worst case of 38mm, is limited by the LSB truncation. The maximum distance error is at the closest range at 3.53cm.

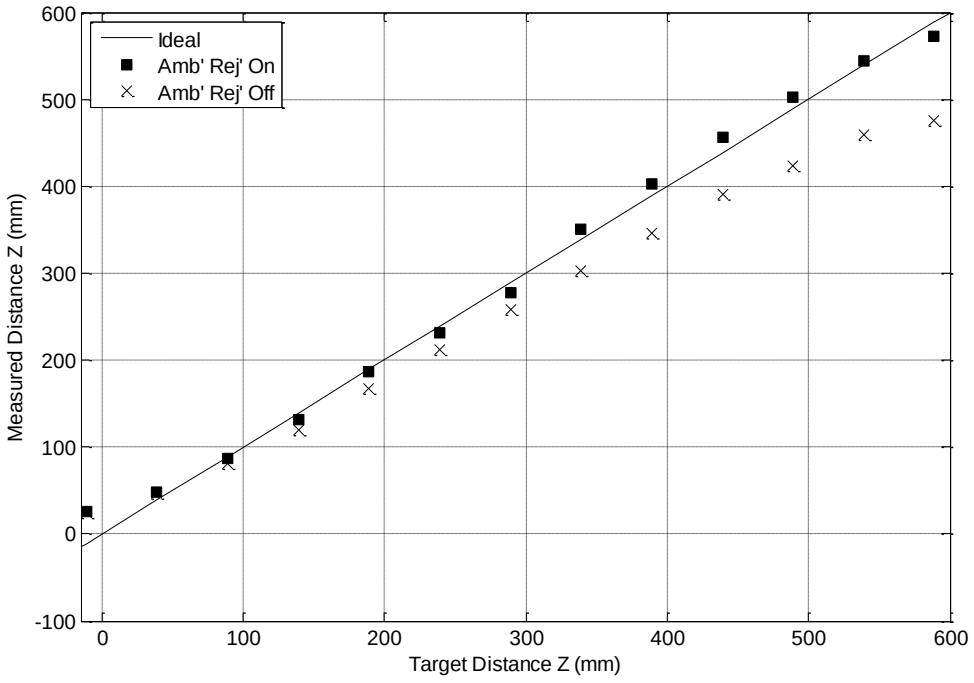


Figure 6.54. ITOF range sweep for 4ns laser.

The ideal precision for an ITOF counter based system in relation to the maximum differential signal is shown in Figure 6.56. The maximum full well of the pixel must be above 1,000 SPAD events to achieve less than 1cm error for pulse lengths <5ns. Ambient light increases the time that a digital oversampled ITOF system must integrate for, to maximise the signal collection from the received laser echo.

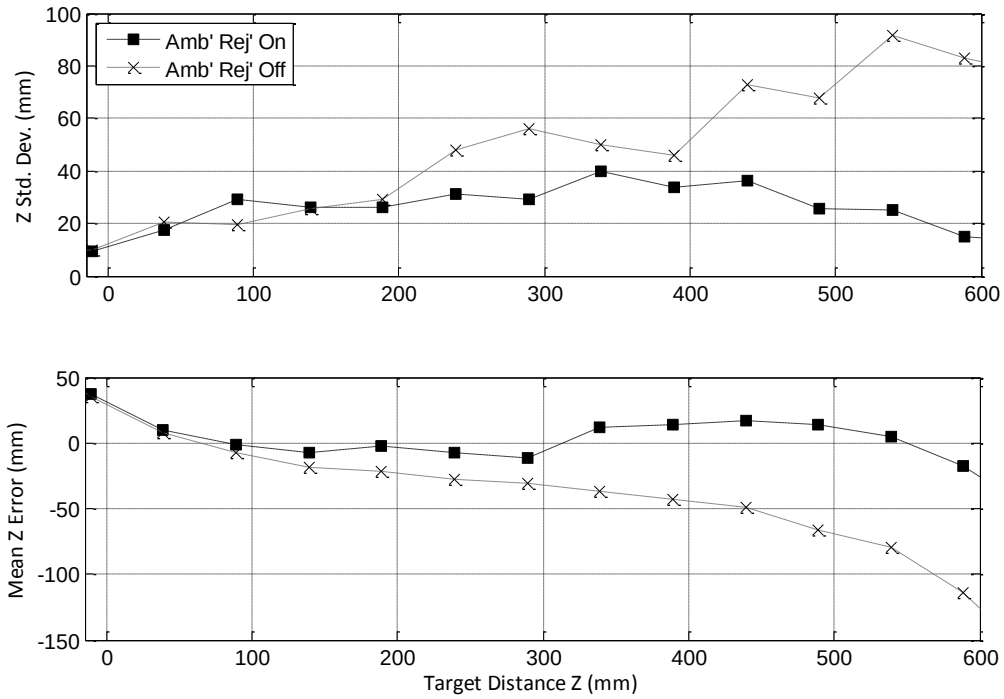


Figure 6.55. TOF precision and accuracy for a 4ns laser.

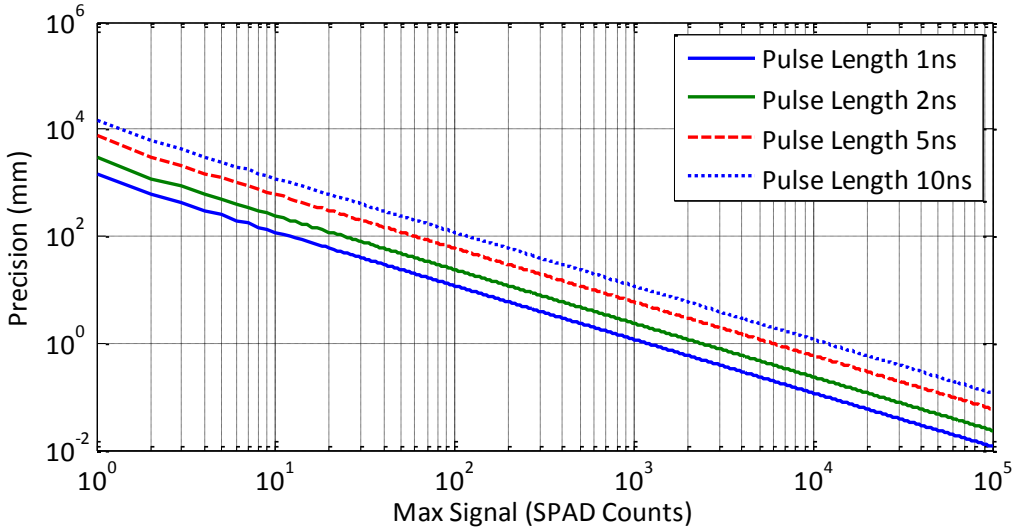


Figure 6.56. Ideal TOF precision versus maximum integrated differential signal.

6.6.2. Digital Oversampled ITOF Images

Figure 6.57 illustrates the modified experimental setup used to capture ITOF images of a toy placed on white paper in front of a white board. The sensor is set, with four bin ambient subtraction, to acquire 2048 bits for each of the four windows with 128 laser repetitions per bit. The total exposure time is 153.6µs in a 4.1ms acquisition time (removing the time of USB transfer overhead). The sensor is in portrait orientation and is fitted with a F#1.4 lens. No optical notch filter is used.

Figure 6.58 demonstrates an example oversampled frame image of each of the two bins 'A' and 'B'. The toy has similar illumination in both images, whereas the background is noticeably darker in the bin 'A' image and the dark shading at the edges is due to the time-gate mismatch across the sensor.

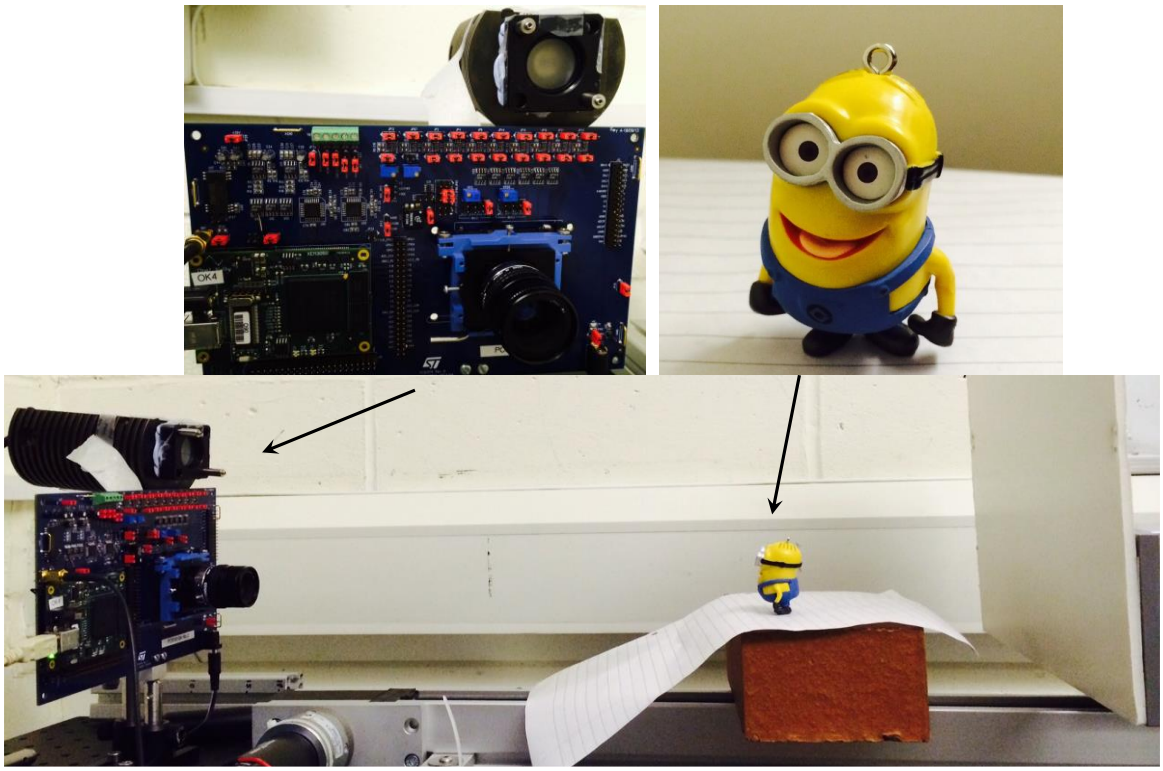
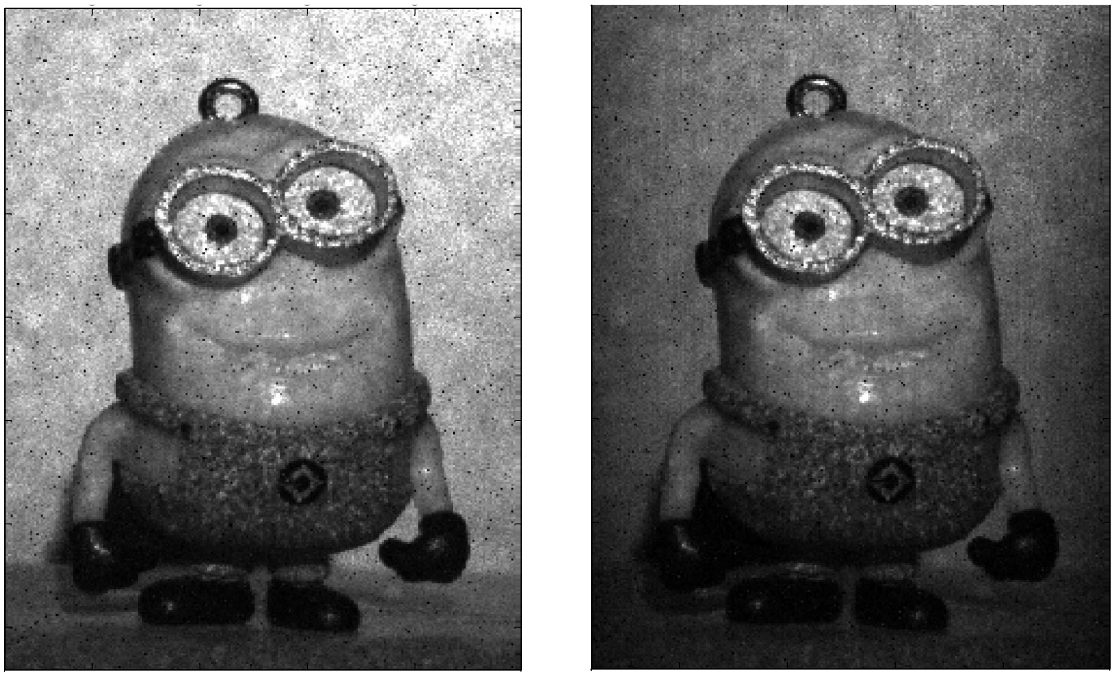


Figure 6.57. Experimental setup for capture of ITOF images.



(a) Bin B Image (b) Bin A Image

Figure 6.58. Individual bin images used to compute the TOF image.

The top left image of Figure 6.59 illustrates the computed uncorrected TOF image, the time-gate mismatch is evident in background which gives the appearance of a bowed background. The top right image shows the effect of applying a spatial median filter with 2x2 kernel which suppresses range noise. The bottom pair of images demonstrate the effect of applying the threshold algorithm detailed in Section 5.4.6 which removes low intensity regions which have high range noise (seen in top left images in the hands and feet of the toy) from the final image. The bottom right image has both a 2x2 median filter and threshold algorithm applied.

Figure 6.60 illustrates the TOF images with and without median filtering in a 3D rendered view. The upper pair of images have the intensity of the 'Bin B' image applied as an overlay. In the bottom pair of images, the falling edge mismatch is again evident in the bow of the background target.

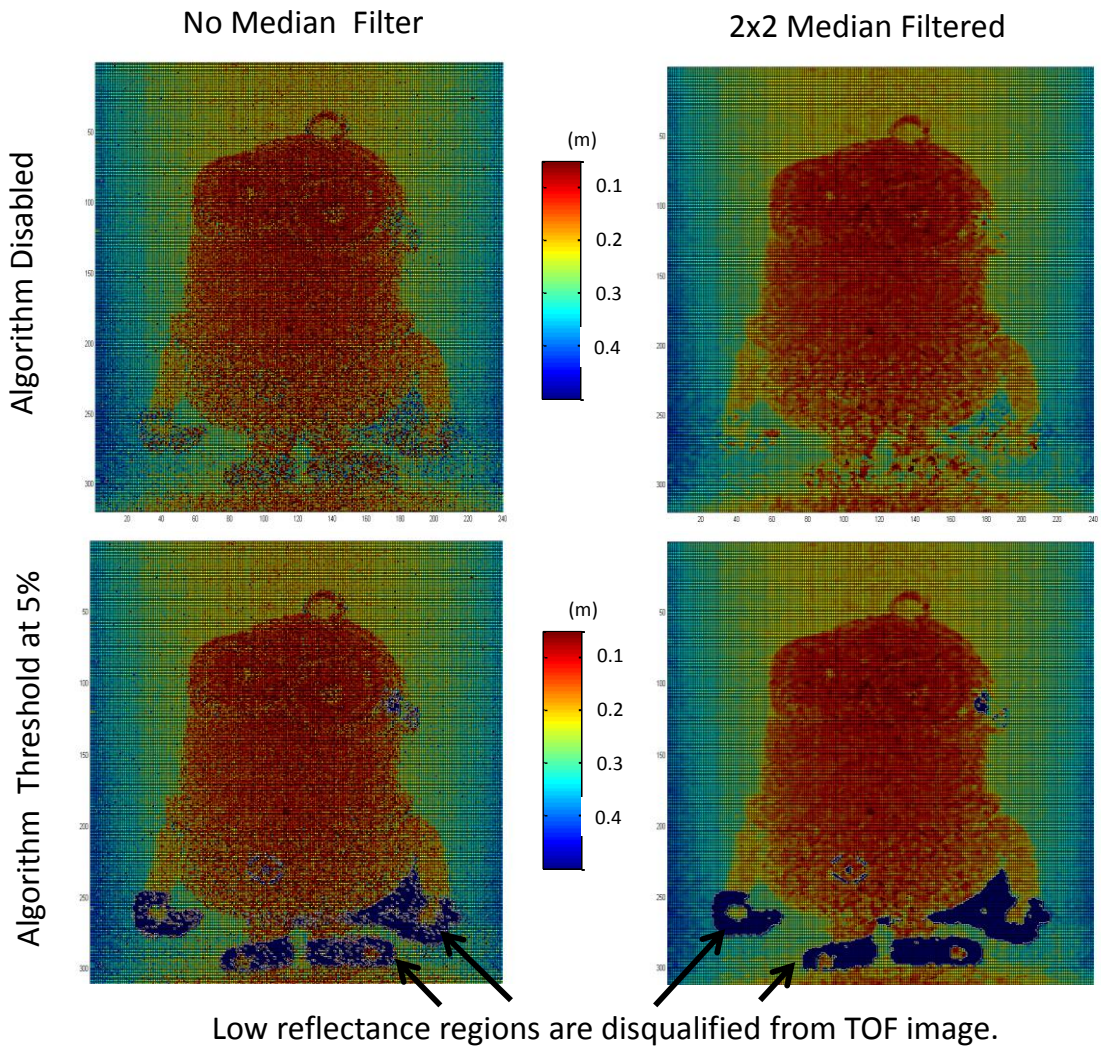


Figure 6.59. TOF image algorithm comparison. The images are coloured in depth and the scale is shown beside the images.

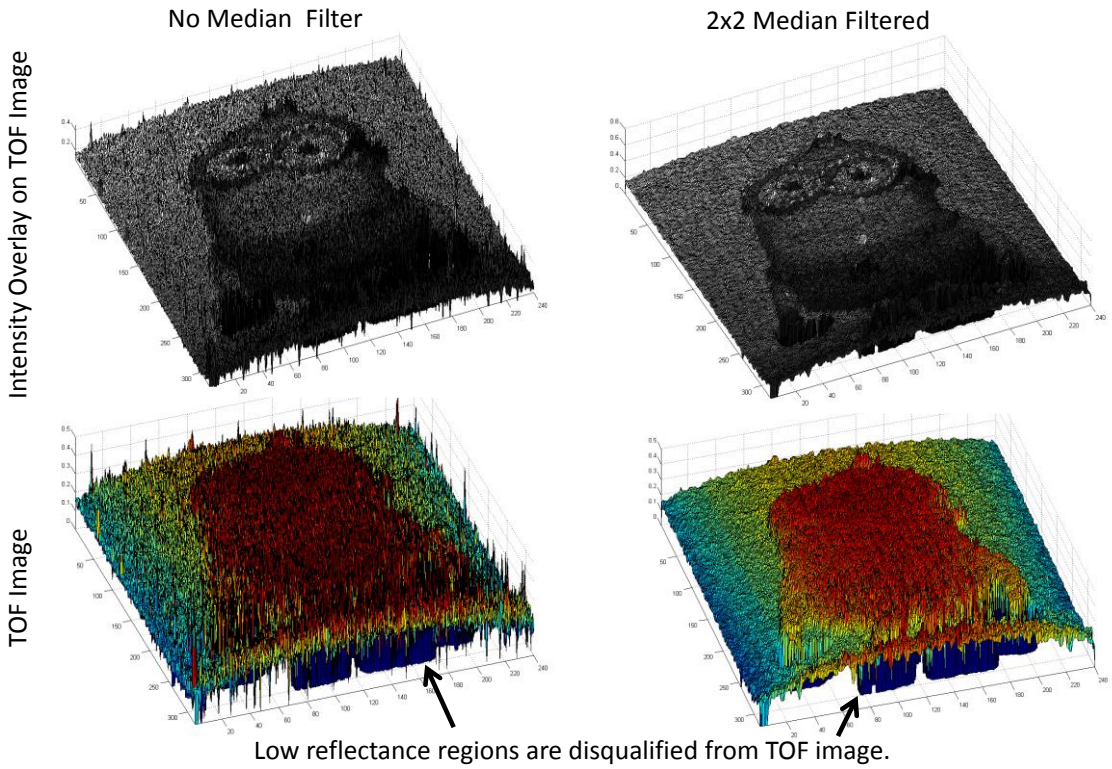


Figure 6.60. 3D rendered view of the TOF depth maps. The upper pair of images have an intensity overlay.

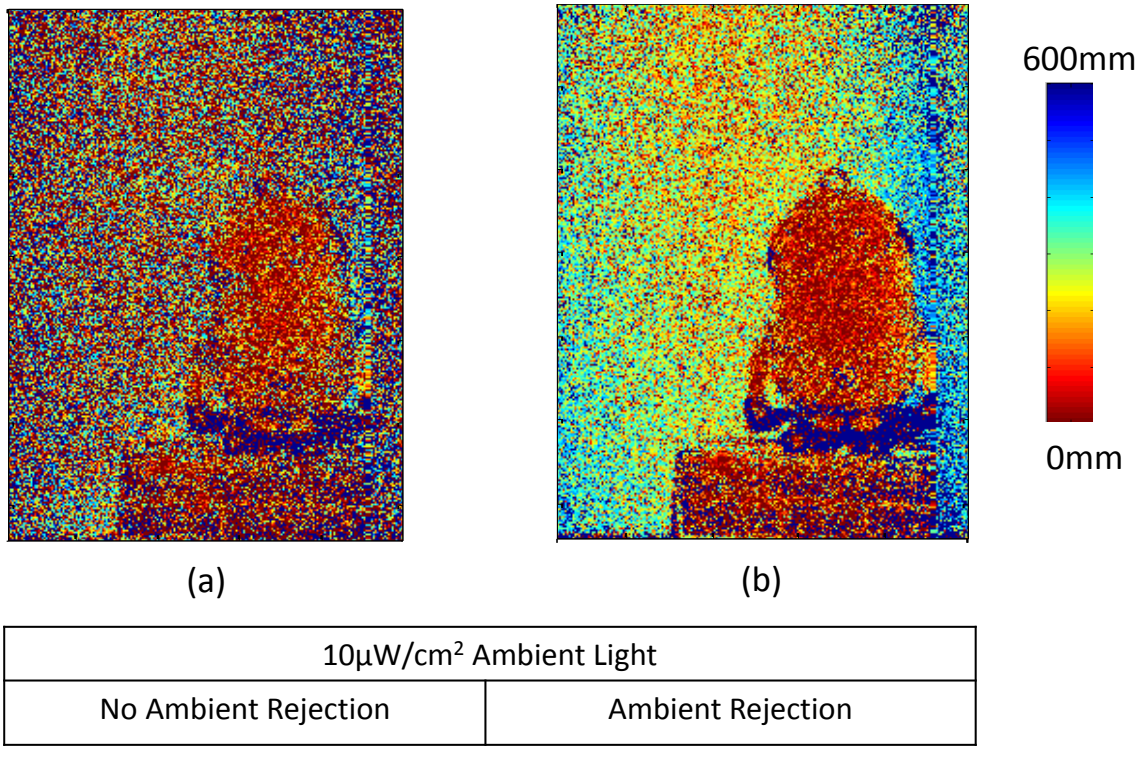


Figure 6.61. TOF depth map shown without ambient rejection in (a), and with rejection in (b). F#1.4 lens employed.

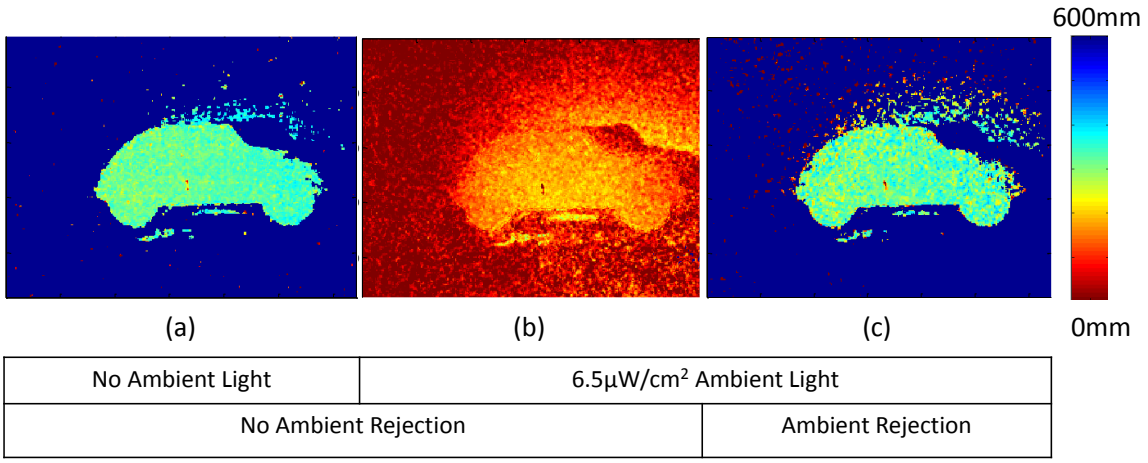


Figure 6.62. TOF images of a toy car at 30cm depth, with and without ambient rejection. Distance range is colour coded. F#2.0 lens in use.

To compare the effectiveness of the ambient rejection method, side by side images are shown in the next two figures. An 850nm optical notch filter (40nm bandwidth) is added to the experimental setup. Figure 6.61 demonstrates two TOF frame images in 10µW/cm2 ambient light measured at the sensor using a ThorLabs optical power meter after the optical notch filter. Figure 6.61(a) is captured only counting up without ambient rejection with many pixels saturating and (b) uses the count up and down method. Distance noise remains evident in the background of image (b).

Figure 6.62 shows a similar image (published in [6]) captured with a F#2.0 lens. Image (b) shows the corruption of the depth map with the addition of the ambient light. The ambient rejection enabled in (c) is successful at retaining the original image in (a). In image (c), the threshold algorithm is employed but additional distance noise is evident around the edge of the car.

6.7. Summary and Conclusions

This chapter details experimental results of the high spatial resolution SPAD-based image sensor. This section provides a summary of the image sensor in each of the different modes. Single photon sensitive imaging is achieved in two distinct imaging modalities: with a multiple photon full well employing analogue counting, and in a binary QIS mode with single photon full well at kFPS frame rate. Digitally oversampled ITOF imaging demonstrates the time gating performance of the sensor. Table 6.16 summarises the general performance data, of the image sensor, provided in this chapter.

Parameter	Min	Typ.	Max	Comment/Units
Process		130nm		LV Imaging CMOS
Chip Size		10.54		mm^2
Spatial Array		320x240		Pixels
Pixel Pitch		8		μm
Fill Factor		26.8		%
PDP (at 480nm)	21.2 (1V EB)	32.4 (2V EB)	39.4 (3V EB)	%
PDE (at 480nm)	5.68 (1V EB)	8.68 (2V EB)	10.6 (3V EB)	%
Median Dark Count	55.8 (1.5V EB)	346.3 (2.5V EB)	1728.4 (3.5V EB)	CPS
Dead Time (1V EB)	1.1	10		ns
SPAD Jitter		184		ps
Power Consumption (Digital)		5.72		mW
Sensor Power Consumption (Analogue)		6.12		mW
SPAD Array Power Consumption	35.5*		610	mW
Sensor Clock Rate	5	48	96	MHz
Frame Rate (Digital)		10	20	kFPS
Frame Rate (Analogue)			6	FPS

Table 6.16. SPC Imager 'AB' Performance Data. * 'AA' sensor revision before DCR reduction.

The sensor and CTA array power are scalable to megapixel arrays, assuming a linear scaling to 1Mpixel would produce <100mW consumption. On the other hand, the SPAD array power consumption for a 1M array is a concern, from 0.5W (DCR limited) to 8W in high light levels. Such a sensor would be best suited to low-light scientific imaging and would need temperature cooling and stabilisation to reduce the DCR. If cooled, the total power of 1Mpixel imager could feasibly be below 200mW for low-light applications.

Table 6.17 contrasts recent monolithic-CMOS works attaining (or close to attaining) single photon counting operation. The first two works are PIN photo-diode based and have heavily optimised the pixel CVF. Ma and Fossum's work is the first to demonstrate true SPC without avalanche gain [58]. The 1.4μm pixel size will permit many mega-pixel SPC imager arrays to be created in the future. In Tohoku University's recent work [167], they have a lateral overflow integration capacitor (LOFIC) which allows CVF and full well to be determined separately (hence the large 200ke- full well). The 0.41e- read noise is close to attaining SPC, and will fall further as this team improve their CVF in future works.

As shown in the table, a SPAD combined with an analogue counter has the benefit of high conversion gain but the disadvantage of cumulative noise. The full well of the image sensor was less than the Chapter 4 test array, which was determined with the FPGA pulse test. For true SPC with photon shot noise limited operation, the image sensor is limited up to approximately 8 SPAD events. The full well of the image sensor can exceed approximately 80 SPAD events, but in this region of operation discrete SPC peaks are not discernible with cumulative CTA noise dominant.

The EMCCD has similar pixel size and full well (at maximum gain) to the SPADs with analogue counters. Like the bias controlled sensitivity of the CTA, the full well of the EMCCD is adjustable by the EM gain. Without EM gain, the read noise is in the order of 50e- but reduces below 1e- after EM gain [168]. The excess noise factor is often cited as problematic for EMCCD, but flattens out to 2x with multiplication gain >10 [43]. The EMCCD requires cooling to -100°C for operation, where the other works in the table operate at room temperature.

Author Surname	**Tohoku**	**Ma, Fossum**	**Andor**	**Perenzoni**	**This Work**	
Reference	[167]	[58]	[168]	[149]	Chapter 4	Chapter 6
Photodetector	PD + LOFIC	'Pump-gate Jot' PD	EMCCD	SPAD	SPAD	SPAD
Array Size	1280x960	1	128x128 to 1024x1024	160x120	3x3	320x240
Pixel Size (μm)	5.6	1.4	24 to 13.3	15	9.8	8
Fill Factor (%)	30.4	-	100	21	3.12	26.8
Pixel CVF or Equivalent	240μV/e-	426μV/e-	Gain dependent	16.5mV/ SPAD event	13.1 to 2mV /SPAD Event	17.4mV to 8.4mV /SPAD event
Full Well or Equivalent	200ke-	210e-	160ke- to 160e-	41	80 to 360	56 to 79
Read Noise (or Equivalent)	0.41e-	0.28e-	<1e- †	0.08e-	0.03e- to 0.22e-	0.06e- to 0.11e-
Cumulative Noise	N	N	*	Y	Y	Y

Table 6.17. Single photon counting image sensor comparison table. * Excess noise from multiplication process. † With EM gain.

Although not added to the table due to technological prematurity, two alternative pixel topologies show promise for in-pixel gain for photon counting. An in-pixel two stage EM gain structure was demonstrated recently under the 'EMCMOS' banner [169]. Also a pinned APD in a standard 4T pixel permits up to 10x gain in-pixel [170].

EBCCD and EBCMOS have not been included in this table, as all these works do not have (or may be combined with) external photocathode image intensifiers. An EB intensifier system coupled with a cooled BSI-version of this SPAD-based image sensor (with no metallisation over the imaging array to avoid interference with the electron bombardment) would yield a low light camera with unmatched low light sensitivity.

The two main sources of cumulative noise are kT/C and temporal noise. Several future improvements may be made to lessen the input referred noise and cumulative noise using the CTA pixel in single photon counting applications:

- Optimisation of the reset row driver should be investigated, the driver's NMOS pull-down is oversized and will contribute charge injection noise on the pixel capacitor.
- On-chip supply decoupling must be increased to reduce high frequency temporal noise in the row drivers, CTA bias lines, pixel and readout.
- Adding a row-wise regulator for the source bias voltage (or per several rows) to ensure supply stability, supress temporal noise and to reduce supply fluctuations.
- Implement a column parallel programmable gain amplifier, with a band-pass filter, in the column readout to increase the signal to readout noise ratio (op-amps and ADC temporal noise contributions).
- Applying a second capacitor in-pixel to store the reset level would permit true CDS and mitigate the kT/C reset noise contribution.
- Increase the size of both the CTA parasitic and main capacitors to maintain the capacitor ratio but reduce the kT/C contribution of the cumulative CTA noise.

Oversampled binary or quanta imaging is an emerging field of research. Table 6.18 provides two QISs compared to this work. Masoodian et al. have recently presented the first photo-diode based QIS, attaining a similar data rate to this doctoral research (lower frame rate but greater number of pixels). The disadvantage of their work is the single row rolling shutter exposure, which is not a photon efficient scheme and not suitable for scientific imaging. Although their recent test structure [58], cited in the previous table, is a 4T structure which will allow conventional multiple row rolling shutter integration.

All three works employ NMOS only pixels, but Burri et al. make use of NMOS-only SRAM with a static bias current which can be seen in the high power consumption of the device. The frame rate in their work is an order of magnitude higher by greater number of parallel data outputs. The high power consumption is reflected in the power figure of merit (FOM). The SPC Imager has a lower FOM for the full sensor than Masoodian et al., but did not have the option on-chip to measure the power consumption of only the front end comparators for comparison. Neither Masoodian et al. or Burri et al. describe measurement of the BER or equivalent read noise. For future QIS sensors, giga-FPS frame rate will only be achieved through on-chip oversampling and low voltage differential data interfaces for off-chip transmission.

Author Surname	**Masoodian, Fossum et al.**	**Burri, Charbon et al.**	**This Work**
Reference	[171]	[139]	[2], [17]
Sensor Name	QIS Pathfinder	SwissSPAD	SPC Imager
Process Technology	180nm CMOS	0.35µm HV CMOS	130nm Imaging CMOS
Array Size	1376x768	512x128	320 x 240
Photo-detector	'Pump-gate Jot' PD	SPAD	SPAD
Fill Factor (%)	45	5	26.8
Pixel Pitch (µm)	3.6	24	8
Microlensing	N	Y	N
Shuttering	Rolling	Global	Global
CDS	True CDS	None	None
Parallel Output Data Channels	32	128	16
Max Frame Rate (FPS)	1,000	150,000	20,000
Max Data Rate	1Gbps	10.24Gbps	1.54Gbps
Pixel CVF or Equivalent	120µV / e-	>1V per SPAD Event	>1V per SPAD Event
Bit Error Rate	Not Reported	Not Reported	1.7×10^{-3}
Read Noise (e- or equivalent)	Not Reported	Not Reported	0.168e-
Power during operation	20mW	1650mW	5.72mW (Sensor) + 35.1mW (SPAD DCR)
Power FOM†	2.5pJ/b (ADC only) 19pJ/b (Full Sensor)	168pJ/b (Full Sensor + SPADs)	14.5pJ/b (Full Sensor) 104pJ/b (Full Sensor + SPAD DCR)

Table 6.18. Quanta imaging comparison table. †FOM=Sensor power/ (No. of Pixel x FPS x N), N = ADC resolution (1b for these sensors).

The TOF performance is compared against recent works in Table 6.19 using the table format from [35]. This sensor attains the smallest pixel pitch of the compared works and microlensing may be employed to increase the optical fill factor further. The relative precision of this work is unfavourably high in comparison due to the low illumination power, and 8 bit depth of the counters. No correction is applied for the systematic spatial variation (or shape) of the time gating falling edge. Furthermore, no correction is applied for the logarithmic 'D-log-H' exposure characteristic whilst using the linear TOF equation which will induce greater repeatable depth error.

Author Surname	**Niclass et al.**	**Stoppa et al.**	**Walker et al.**	**Kim et al.**	**Niclass et al.**	**Bamji et al.**	**This Work**
Reference	[122]	[52], [172]	[39], [53]	[173]	[35]	[174]	[6]
Process	0.35µm HV	180nm Imaging	130nm Imaging	110nm Imaging	180nm HV	130nm Imaging	130nm Imaging
Sensor Identifier	-	-	Phase Domain ΣΔ	RGBZ	Toyota LIDAR	Kinect 2	SPC Imager
Array Size	60x48	80x60	128x96	480x270	340x96	512x424	320x240
Photo-detector	SPAD	Lock In PD Pixel	SPAD	Lock In PD Pixel	SPAD	2-tap Photogate	SPAD
Fill Factor (%)	<1	24	3.1	34.5	70	60	26.8
Pixel Pitch (µm)	85	10	44.65	14.6*	25	10	8
Microlensed	N	N	N	Not reported	N	Y	N
Frame Rate (FPS)	22	5	20	11	10	30	4.6
TOF	Indirect	Indirect	Indirect	Indirect	Direct	Indirect	Indirect
In-pixel Bins	2	2	2	2	N/A	2	1
Illuminator Repetition Rate or Frequency (MHz)	30	20	3.33	20	0.2	10-130	22.5
Illuminator Signal Power (mW)	800	80	50	55†	40	Not reported	20
Distance Range (Measured)	2.4	6m	2.4m	4.5m	100m	0.8 to 4.2m	0 to 0.6m
Absolute Precision (cm)	3.8	16	16	3.8	<10	1.5	3.8
Relative Precision (%)	1.6	2.7	6.7	0.84	0.1	<0.5	6.3
Worst Case Distance Error (cm)	11	4	0.5	4.2	36.6	2	3.53

Table 6.19. TOF imaging comparison table. * 4x4 binning of 3.65µm employed in this work equalling 14.6µm pitch. † Peak LED power attained from part datasheet.

7. Summary, Conclusion and Outlook

7.1. Summary

This thesis set out to verify that CMOS SPADs could be efficiently close-packed beside analogue pixel electronics to form a scalable foundation to build a high spatial resolution time-resolved SPAD-based image sensor for single photon counting and time-of-flight imaging. The research presented, illustrates the design, development and effective operation of the SPAD-based QVGA imager in both these applications.

Chapters 2 and 3 provides a background overview to SPAD operation and construction. A review and study of different SPAD diodes and guard ring structures indicates that the Richardson P-well to deep N-well SPAD is favourable to other structures for PDP, DCR and ease of integration. This substrate isolated structure is compared, to the Webster deep N-well to P- substrate (non-substrate isolated) SPAD, in terms of scaling to pixel pitches below 10μm indicating that the former is both a truly scalable and compact structure achieved through well sharing techniques.

Chapter 3 presents an overview of the plethora of SPAD-based pixels grouped by circuit architecture. As revealed in the review of the state of the art, and explored in Chapter 4, analogue pixel circuits are the route in monolithic CMOS, to low pixel pitch, high spatial resolution SPAD image sensors. It is determined that for analogue counters, the CTA architecture provides lower non-uniformity than SCS designs.

Chapter 4 presents and discusses the single photon counting performance of a CTA-based SPAD pixel test array. Non-ideal SPC behaviour is analysed and attributed to a combination of cumulative noise per SPAD event and a distortion unique to CTA analogue counters. Nonetheless, SPC is demonstrated with input referred noise below the equivalent of 0.15e- with a bias controllable full well from around 80 to 300 SPAD events.

The image sensor architecture and design is described in Chapter 5. Simple yet effective time gating, single bit binary readout, analogue CDS readout and row driver logic are shown and the four modes of image capture are described. Intensity imaging is achieved by both global shutter analogue photon counting and oversampled single photon binary imaging. Time gated image capture is similarly described to realise oversampled digital TOF imaging.

The compact and local well sharing SPAD design is measured and evaluated achieving comparable performance to previous non-shared structures. The pixel achieves the smallest SPAD pixel pitch and highest fill factor of published SPAD-based image sensors to date (without microlensing). The analogue photon counting is

effective at capturing single photon counting images at sub-0.3e- equivalent noise up to 5 photons, and 1e- noise up to 19 photons. Single photon binary field images are oversampled in FPGA to form multi-bit frame images, realising a SPAD-based QIS. Theoretical QIS performance is confirmed experimentally, bearing close resemblance to the light to exposure relationship of photographic film.

The time gated imaging capability of the image sensor is characterised in Chapter 6. Two fundamental design issues preclude capturing gated analogue SPC imaging. First, power supply drops cause >10 nanoseconds of skew in the time gate leading edge, and less than 1ns mismatch in the falling edge. Secondly, the different routing and control of the time-gate enable and disable signalling creates difficulty in rendering accurate time-domain imaging around the occurrence of the falling edge. Despite this, digitally oversampled TOF imaging is successfully demonstrated.

7.2. Conclusion

This section discusses the merits and demerits of the SPAD design, the pixel electronics, and the time gating in achieving the research project goals.

7.2.1. SPAD Design

The research presented in this thesis demonstrates that the SPAD with local well sharing combined with analogue pixel electronics is currently the most efficacious method to achieve a small pixel pitch and high fill factor SPAD-based imager pixel. Yet more than 50% of the area of the 8µm SPAD pixel remains dedicated to the SPAD guard ring and the spacing of the SPAD guard ring from the pixel electronics. This issue will fundamentally remain for all monolithic SPAD pixels requiring high fill factor, regardless of the process technology. Moreover, to achieve higher spatial resolution SPAD image sensors shrinking the pixel is a must. The pitch could feasibly reduce down to 3 to 5µm, yet maintaining the fill factor above 10% will be challenging. One further disadvantage of the dual strip well sharing technique is the different spatial resolution and MTF in X and Y, but microlensing may partially address this.

Chip stacking technology is the greatest technological leap for image sensor design since the move from CCD to CMOS. Leveraging this for SPAD image sensors will permit either pixel pitch below 3 to 5µm for an analogue counting solution, or the ability to capture TOF in-pixel with digital pixel logic below 10 µm. Improving the limited sensitivity of substrate isolated SPADs for NIR imaging, stacking technology allows the creation of optimised SPAD structures, with improved NIR QE for instance, by customising the top sensing layer. Employing deep trench isolation, increased epitaxy depth and buried oxides as reflectors will improve sensitivity and optical crosstalk, whilst intrinsic or extrinsic drift fields will improve the time-domain performance. The use of backside illumination (BSI) will allow the fill factor to reach 100%.

7.2.2. Pixel Design

An analogue counter is shown in this thesis to be effective for single photon counting. Accurate counting is demonstrated for the first few photons with sub 0.3e- read noise. In comparison to other imaging technologies, the conversion gain and effective full well are comparable to EMCCD technology. On the other hand in comparison to sCMOS, the full well is one to two orders of magnitude lower, but the read noise is also an order of magnitude lower for the first few photons.

Although not the intention of this research, the developed SPAD image sensor would be suitable for global shutter machine vision and security applications. With optimised timing and the addition of column parallel ADCs, high frame rates could be achievable and commensurate with global shutter image sensors. For a commercial imaging product, a full well of 100's of photons (or SPAD events) requires oversampling. This has the significant disadvantage of requiring both an external frame store and data processing, which are not suitable for many applications.

For a TOF 3D vision sensor product using SPADs, to be competitive with 2 or 4-tap PMDs, the single counter design may be extended to a pixel containing 2 or 4 counters. Conceivably either an all-NMOS pixel with four CTAs and time-gates (count down only), or a two bin pixel with PMOS for counting both up and down. With the downside that adding greater numbers of transistors would decrease fill factor. A counting capacity of >10,000 counts per pixel is an ideal specification as it matches the full well capacities of PMDs. However, using analogue counting in a 1V range, this gives a single step voltage of 100μV. As demonstrated in Chapter 4, step-sizes below the millivolt range are subject the counter to large non-uniformities and noise accumulation. Even with design optimisation to minimise the Vt variation of the CTA counter and minimisation of accumulated noise, the maximum counting capability of the pixel will remain limited to a few hundred counts for acceptable noise and PRNU. No analogue counter can escape cumulatively adding noise per SPAD event but can be minimised by decoupling supplies and increasing the capacitance on the parasitic node (to reduce the kT/C noise contribution) to the detriment of the maximum counting capability. Three possible solutions are proposed, taking steps towards improving the full well:

- In an all-NMOS CTA architecture, boosting the maximum voltage of the counter capacitance would increase the maximum counting range.
- If a SCS counter is employed, this can be achieved by maximising the capacitor size.
- Alternatively an active CMOS integrator architecture (at the expense of PMOS devices) would yield a large full well by increasing the voltage range of the integration capacitor beyond 2V.

The imager serves as a demonstration and working proof of concept of the QIS concept using the analogue counter in a high gain SCS mode. The source degradation bias, intended for controlling CTA sensitivity, has

the significant advantage in mitigating memory leakage by operating the SCS switch in the cut-off region. This feature should be re-used in future developments of SPAD-based dynamic memory pixels. In a lower geometry DSM CMOS node, the analogue counting performance would suffer due to decreased voltage ranges. On the other hand, the QIS dynamic memory would still function and would benefit from lower leakage and decreased power consumption from the reduced supplies. Readout techniques and architectures from static and dynamic memories would benefit the design of future QIS sensors.

7.2.3. Time Gating

The design of the time gating on the image sensor could be improved with several enhancements in the pixel, the row driver, the balanced tree routing and in the signal generation block. Although issues are in each component, these are not isolated issues; in order to improve the time gating imaging performance they must be considered and designed as interlinking components in a time-domain system.

The pixel time-gate switch must be designed in conjunction with the row-driver: time gating any SPAD front end circuit with two signals introduces systematic errors into temporal image capture. The pull-down disable on the sensitive SPAD node creates two problems for analogue counting. The first is the time-gate falling edge 'spike' created by the sample and hold style circuit design. The second is the interaction of the disable and enable appears to vary both the onset and slew of the enable rising edge. This has yet to be investigated fully, but is assumed to stem from the same power supply issue that causes the 'bowing' row driver mismatch. It is apparent that adequate decoupling needs to be included for the row drivers for the next iteration of this image sensor.

Furthermore, the disable needs to be routed through a balanced clock tree. To immediately fix the time gating, the disable row signal should be altered to be a logical negation of the enable signal. Care must be taken to ensure the overlap between these signals is minimised. Ideally these should be non-overlapping signals whereby the disable signal falls just before the enable assertion and rises again just after the enable but this is difficult to realise. Without careful management of the disable to enable signal overlap, the disable transistor becomes the lowest resistance path for the SPAD anode causing unpredictable diode behaviour. In the 'TACIMAGER' IC, both disable and time-gate enable are generated through delay generator blocks which goes some way toward achieving this goal [7].

Such issues would be avoided if the time-gating could be performed using a single input signal applied only to the gate of an NMOS either in line with the counter or modulating the 'V_B' discharge node of the counter (or some similar approach). On the other hand, the single inline switch time-gate, without disable pull down, is ideal for a time-gated binary sensor where the 'latching' of SPAD events into the time-gate input intentionally uses the sample and hold issue, to assist in the discharge. It is noted, that this 'latching' front end circuit has been recently adopted alongside a CTA pixel design by the FBK research group removing the disable transistor

but limiting the sensor operation to only capturing one photon per laser excitation cycle in a time-gated FLIM application [149]. Time gating in the digital domain has advantages over an analogue approach, as the circuit response is only sensitive to the time-gate edge at the crossing point from low to high, and not the whole pulse profile.

7.3. Outlook

This section describes future improvements and enhancements to the sensor and control, and describes taking this research forward beyond this thesis.

7.3.1. QIS Timing Optimisation for ITOF

In an ITOF application, the central limitation of the single bit, single time gate, pixel operation is the limited output frame rate. 3D vision sensors for automotive and consumer applications (such as detection and tracking of users fingers, hands, and bodies) require high frame rates (100's FPS to 1k FPS) with short millisecond acquisition times to successfully image, track and process dynamic changes of a fast moving scene.

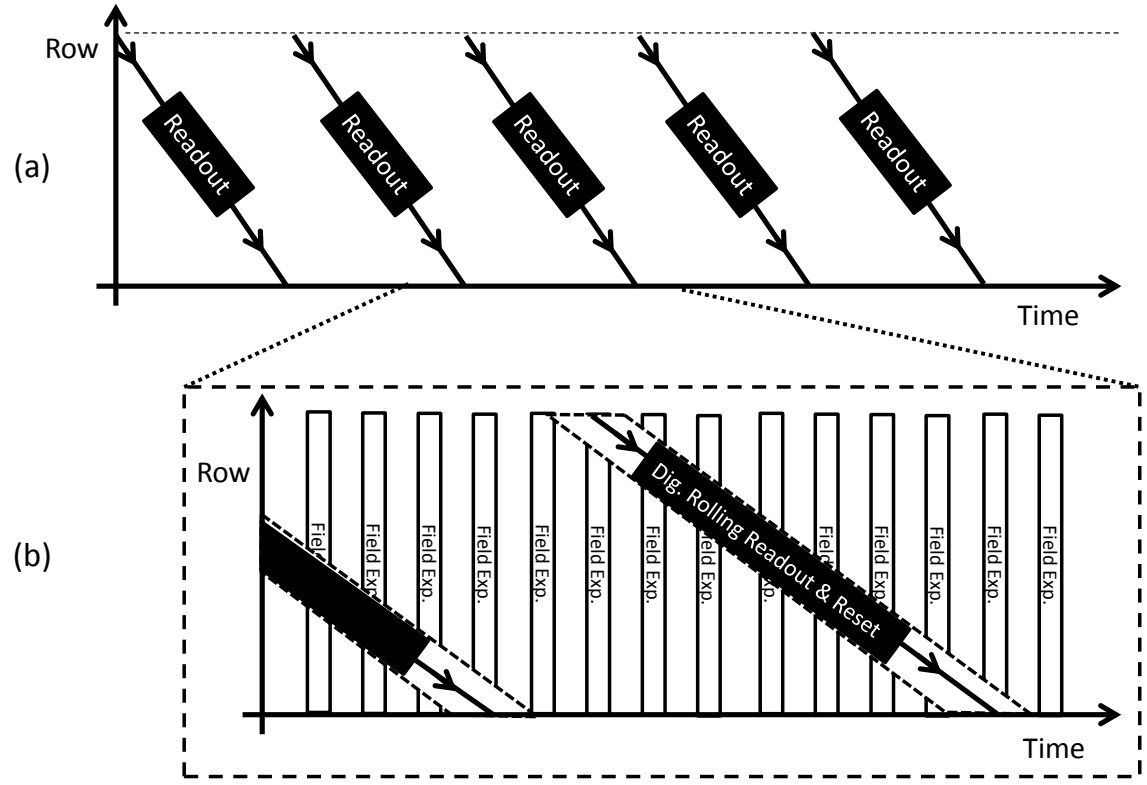


Figure 7.1. Proposed improved sensor readout timing. (a) Diagram showing continuous rolling readout. (b) Highlighted single field exposure is between the two read and resets.

The acquisition time of the current SPAD-based QIS mode is dominated by two factors: non-optimised timing of field image capture, and limited sensor output bandwidth to the oversampling data processing on FPGA. In order to create a high frame rate ITOF sensor, three methods are proposed to address the speed limiting factors and increase the output TOF frame rate.

First the digital single bit mode, like the analogue counter, could be expanded to four memory elements each with an individual time-gate input. This would increase the field image rate by 4. The next proposal is to implement the TOF camera equivalent of a rolling shutter. Figure 7.1 illustrates the improved sensor timing with distributed global shutter exposure through the continuous rolling readout. Every row integrates for 239 of 240 total exposure periods, and reads out and resets during the 240th. Using a 40MHz clock as an example this would take the acquisition time of a field image from 220µs plus exposure time (100µs global reset plus 4800 clock cycles or 120µs for readout) down to the 120µs readout time only. This would effectively halve the existing acquisition time and double the output frame rate. The final proposal addresses the system output data rate. The concept is twofold: maximising the sensor clock rate to minimise the readout time, and employ spatial pixel binning. Binning achieves the necessary signal rate acquisition in the short frame time required by the TOF application. Table 7.1 indicates the advantage gained in output frame rate at the cost of spatial resolution and array size. The assumption is made that a ITOF image is constructed using the four bin capture presented in Chapter 6 and in [6]. A further assumption is made that 5000 counts (20000 field captures) are required to be captured, per bin, per acquisition.

Binning	Array X by Y		Acquisition Time for 20k Fields (ms)	Output Frame Rate (FPS)	Frame Rate Improvement
1x1	320	240	2400	0.42	**1x**
2x2	160	120	600	1.67	**4x**
4x4	80	60	150	6.67	**16x**
8x8	40	30	38	26.67	**64x**
16x16	20	15	9	106.67	**256x**
20x20	16	12	6	166.67	**400x**
32x32	10	7	2	426.67	**1024x**

Table 7.1. Spatial binning to increase output frame rate of the ITOF image stream.

7.3.2. QIS Bit-Plane Processing Techniques for High-Speed Imaging and Super-Resolution Microscopy

This section presents an extension to this doctoral research that is underway, exploring using the image sensor for high-speed imaging in super-resolution microscopy. Prof. Robert Henderson's European Research Council (ERC) 'TotalPhoton' grant seeks to develop SPAD-based image sensor technology for super resolution imaging wide-field microscopy to obtain fluorescence images with spatial resolution under the diffraction limit using computational localisation methods. In collaboration with Dr Istvan Gyongy and Luca Parmesan, the single-bit QIS mode of the 'SPCImager' is used as a proof of principle to demonstrate the low-light and high speed imaging capabilities of SPAD-based QISs, and to explore the processing techniques and architectures

required to handle the huge dataset of a live streaming kFPS QIS. In Chapter 6 and [2], conventional temporal oversampling is utilised: for every doubling of frame image bit depth, the output frame rate is halved. In [8], instead a rolling average of bit-plane field images is proposed, to construct an output frame image stream operating at the same frame rate as the sensor. This high speed imaging technique facilitates capturing fast moving events. Figure 7.2 exemplifies this processing technique applied to a 4kFPS video stream capturing the fall of polystyrene beads onto a surface. The upper image sequence comprises consecutive raw bit-planes with 2μs exposure. The lower sequence is a rolling 2-bit sum with increased bit depth and preserved frame rate.

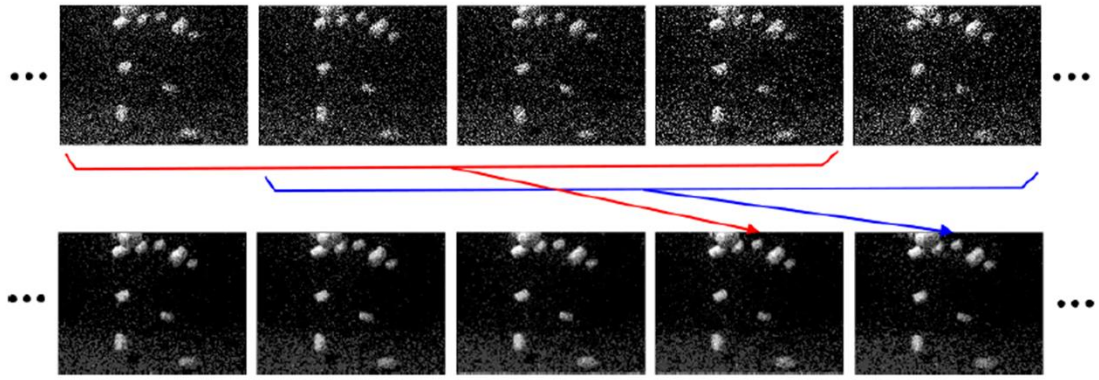


Figure 7.2. Reproduced from [8], courtesy of Dr I. Gyongy. Rolling summation of QIS bit planes showing falling polystyrene beads. Upper sequence are raw bit planes. Lower sequence shows 4 field oversampled, rolling summation 2b frames.

An implementation of the rolling summation technique is proposed that is suitable for FPGA operation and is illustrated in Figure 7.3(a). An extension of the technique is described in Figure 7.3(b), targeted for event-driven readout for photo-activated localisation microscopy (PALM) imaging. PALM imaging is a computational super-resolution technique designed to detect photo-activated 'blinking' of macro-molecules in the order of microseconds. The premise of the proposed technique is to threshold the rolling sum and track the length of time a pixel has been above the intensity threshold. Once a pixel falls back below the threshold, the pixel address, the stop time-stamp and the measured length of time are output, hence identifying a 'blinking' macro-molecule. This heavily reduces the sensor output to the upstream computer by performing event-driven run-length encoding on the image stream.

The reader is directed to [8], for dynamic and adaptive bit-plane post processing and scientific imaging comparison of the 'SPCIMAGER' in side-by-side comparison with a Hamamatsu sCMOS image sensor imaging quantum dots.

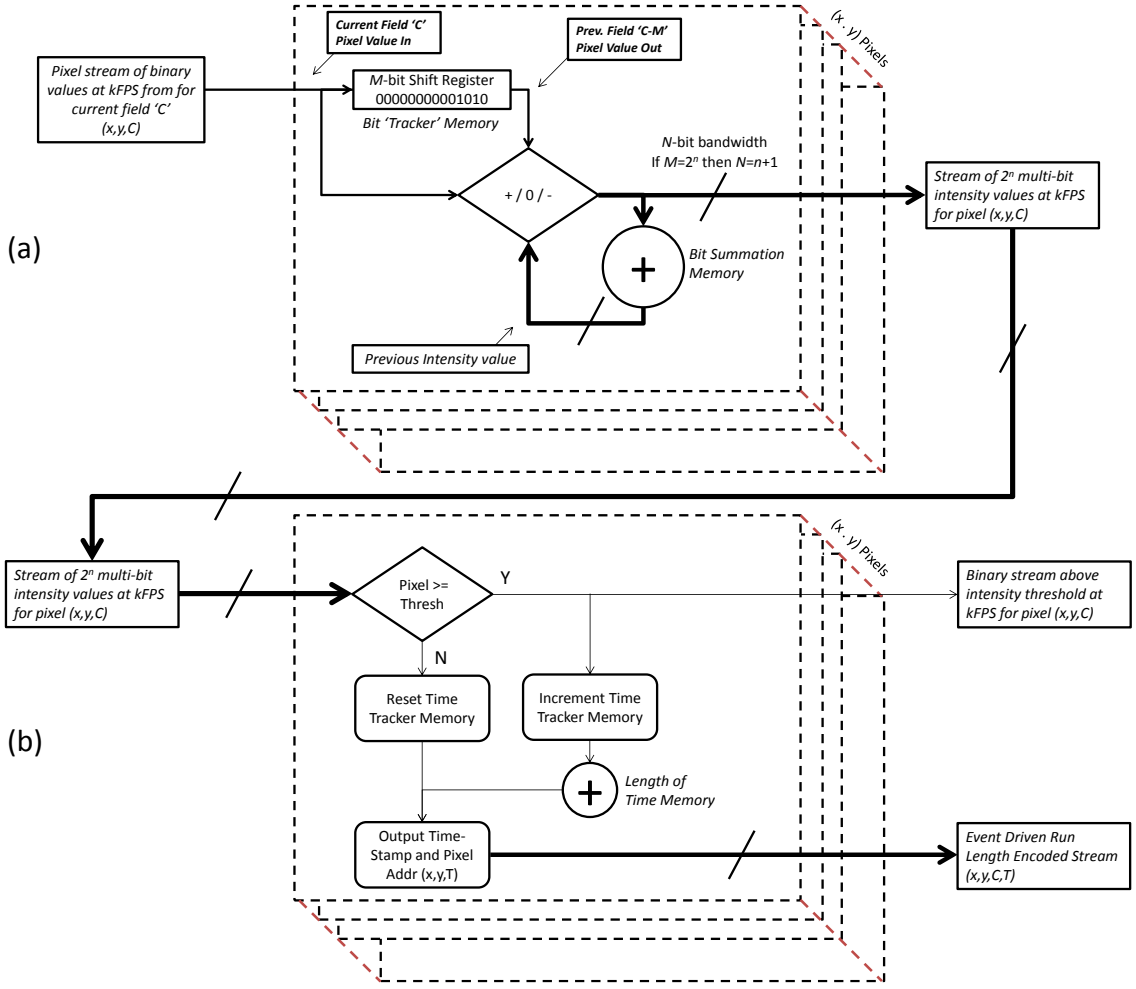


Figure 7.3. (a) FPGA suitable rolling summation technique for processing QIS bit-planes from [8]. (b) Event-driven run-length encoding of the rolling summation for PALM.

Adaptive window oversampling is an extension to the moving averaging technique, altering the shift register length dependent on the variance of the preceding bit planes on a per-pixel basis. Short windows are used for dynamic or changing areas of a scene trading off bit-depth against motion blur. Long windows are applied in slow moving regions creating pixels with high bit depth which do not suffer from motion blur. Appendix 4 has a list of videos captured using the sensor using rolling summation up to 7-bit (128 bit-plane summing) with side by side comparison of conventional, rolling and adaptive rolling temporal oversampling.

The recent work on oversampling by Vogelsang et al. in [175] extended Fossum's original concept of single photon QIS to multi-photon QIS. It is envisaged that the CTA mode of the pixel would permit such multi-photon oversampling. This would have the benefits of preserving single photon sensitivity and large full-well (limited only by the digital memory width used for oversampling) whilst using the high frame rate of the digital readout. A very large dynamic range is conceivably possible, combining the multi-photon counting with an oversampled frame store, which would extend the limited dynamic range of the analogue counter.

7.3.3. Design of the Ideal SPAD TOF Pixel

Weighing up the advantages and disadvantages of the implemented pixel for TOF, this section briefly describes a target for a SPAD-based TOF pixel. The ideal pixel would comprise the following features:

- *An all-digital pixel:* As discussed in Chapter 3, the all-digital SPAD pixel mitigates the contribution of analogue noise and distortion sources. Power consumption becomes a primary concern with high frequency clocks. Utilising a DSM CMOS process with small feature size, and low supply voltages, is the direction to achieve this low power with compact digital pixel electronics. Further into the future, FinFET or FD-SOI processes with geometrical feature size <32nm, are appealing to address this.
- *3D stacked:* Stacking the SPAD above the pixel electronics is the most effective way to gain fill-factor approaching 100% [176]. An all-digital pixel in a lower tier DSM CMOS can be incorporated with an optimised top tier photo-sensing layer.
- *Oversampling:* A method of in pixel time-domain averaging or oversampling should be employed without the requirement of an external frame store.
- *Noise reduction:* Employ an ambient or dark noise reduction scheme to limit the effect of uncorrelated noise. Co-incidence detection circuits warrant investigation (for example the adder tree network in [35]). The system would benefit from improved SNR. Conceivably this may also result in a reduction in time allocated to performing ambient subtraction, which subsequently would increase the maximum frame rate.
- *Ambient light pile-up rejection:* Pile-up distortion in the time converter circuit due to readout or conversion dead time should be minimised to maximise the ability to cope with high intensity ambient light. Converter design strategies such as found in [5] are ideal in this regard.
- *Dual Intensity and TOF Imaging:* The pixel would ideally be capable of global shutter intensity imaging either in a separate image capture mode, or ideally in interleaved intensity and TOF imaging frames to minimise the disparity between the two image streams.
- *Power consumption:* Minimise power consumption using a time converter with a low energy per photon conversion figure of merit.
- *Direct TOF over indirect TOF:* ITOF is inherently limiting, a DTOF design is preferred for numerous application reasons but is inherently challenging to design a DTOF pixel with all the above criteria. If these design challenges can be met then a SPAD-based DTOF pixel would provide many advantages over competitor solutions.

7.4. Final Remarks

The motivation of this research was to create the highest spatial resolution SPAD-based image sensor array yet described in the known literature. This has been achieved through compact SPAD and pixel design. The smallest and highest fill factor SPAD-based image sensor pixel, yet published, has been successfully demonstrated achieving single photon counting and ITOF image capture.

The parallel research, presented in Appendix 2, into a low spatial but high temporal resolution TCSPC sensor design yielded a proof of concept design showing that a TCSPC histogram can be created directly on chip in parallel at GS/s throughput without data compression. This architecture has promise to be expanded and crafted for numerous applications that depend on combined photon-counting and TCSPC such as TOF ranging, PET, fNIRS, cell sorting and techniques such as FLIM-FRET.

It is envisaged that this thesis research will continue in multiple parallel strands; three central developments emerged during this thesis research and will be continued in diverse, and yet aligned, directions.

The most promising application of high frame-rate single photon counting, using the QIS or oversampled binary image sensor mode, is in the field of super-resolution microscopy either though STORM or PALM. The sensor readout and analogue dynamic memory pixel have been re-designed for the 'TotalPhoton' project to increase the frame rate by use of greater parallelisation of data outputs, and push the pixel fill factor beyond 60% through an increased pixel pitch of 16μm which is an exciting development outcome of this work.

High spatial resolution TOF (or time-gated) imaging is best served through digital gating and counting circuits. Investigation into indirect and direct TOF digital pixel design will realise improved real-time SPAD-based TOF and TCSPC imaging. Chip stacking will aid this, but the challenge will remain of the trade-off between incorporating digital logic and pixel size. Finally, the techniques to achieve compact SPAD pixel design are not limited to image sensors, they are also directly applicable to single point, line and large area sensors.

Appendix 1. TAC Pixel Design for TCSPC

In parallel with this thesis research, analogue TAC pixels in a test structure were initially explored by the author and then taken on by Luca Parmesan to form the core of his research into a high resolution TCSPC image sensor. This appendix provides a brief overview of the author's work on these TAC pixels.

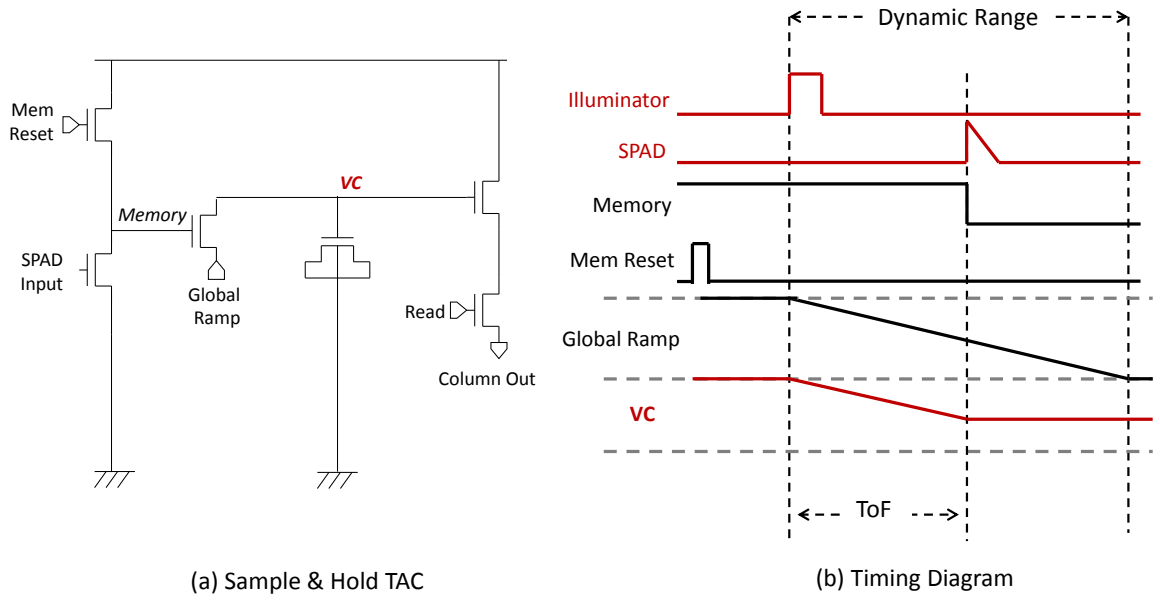
A.1.1. Sample and Hold TAC

A.1.1.1. Test Pixel Structure

On the same MPW as the test structure IC described in Chapter 4, a second IC was developed containing a 3x3 array of sample and hold TAC pixels arrayed in the same 9.8µm pitch with 3.1% fill factor.

The pixel relied upon a basic NMOS sample and hold circuit, shown in Figure A.1.1 (a), triggered by a dynamic memory and integrated into an APS read out circuit. As illustrated in the timing diagram in Figure A.1.1 (b) a laser pulse is activated co-incident with a linear voltage ramp. Upon arrival of a SPAD event the dynamic memory is discharged and the ramp is sampled on the capacitor. The voltage on the capacitor then represents the time of arrival of the SPAD event.

The work was published at ESSDERC 2014 in [4]. At the time of publication, this was the smallest and highest fill factor TCSPC image sensor pixel.



(a) Sample & Hold TAC (b) Timing Diagram

Figure A.1.1. (a) Sample and Hold TAC Circuit Diagram (b) Timing Diagram

A.1.1.2. TCSPC 256 x 256 Image Sensor

The 'SPCImager' formed the building block for Luca Parmesan's doctoral research into a high resolution TCSPC image sensor. To optimise team design effort, the analogue read-out and row decoders were re-used from the 'SPCImager'. Layout and design support was provided by the author. Luca Parmesan re-designed the SPAD and pixel from the test structure to attain 19.8% fill factor in an 8µm pitch. Figure A.1.2 provides a photomicrograph of the 256x256 imager.

The work was published at the IISW conference in 2015 [7]. At the time of publication, this sensor was the highest resolution TCSPC image sensor and contained the smallest and highest fill factor TCSPC pixel in the known literature.

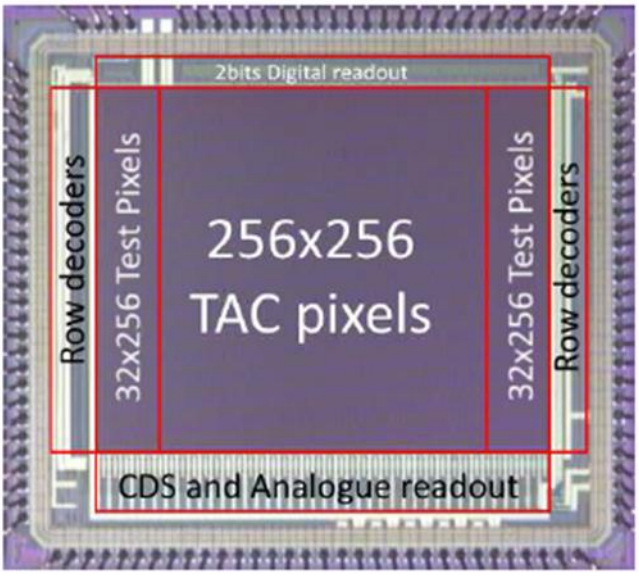


Figure A.1.2. TACIMAGER: a 256x256 Sample and Hold TAC TCSPC-based SPAD Image Sensor

In parallel, Luca Parmesan has investigated a NMOS-only pixel specifically for FLIM with both a SCS counter and integrating TAC. The test structure is a 5x5 array of 10µm pitch pixels at 10% fill factor [150].

A.1.2. Switched Current Source TAC

In parallel with the sample and hold TAC pixel array a third test structure IC was developed to explore Prof. Robert Henderson's proposed circuit for a switched current source TAC. The design was enhanced with a dynamic memory and front-end time gate and the 14T NMOS only pixel schematic is shown in Figure A.1.3. The pixel was designed in the same 9.8µm pitch with 3.1% fill factor. Like the SPC imaging pixel it was designed to have no static bias current. In measurement, the pixel was immediately discounted due the high non-uniformity between pixels in favour of the sample and hold structure. This is similar to high non-uniformity of SCS counter circuits described in Chapters 3 and 4.

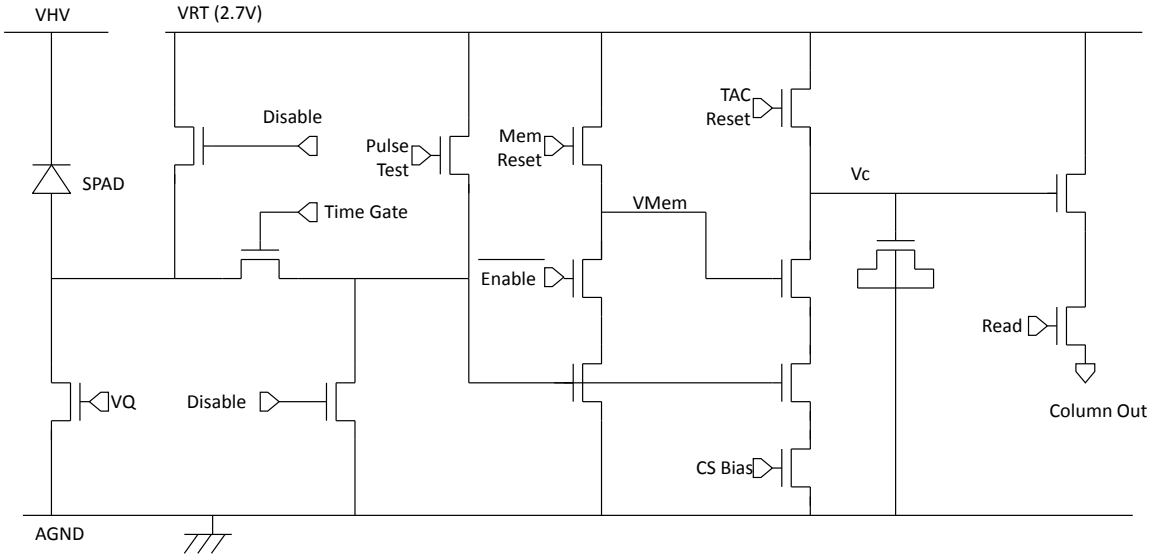


Figure A.1.3. 14T NMOS-only Switched Current Source TAC

A.1.3. Deep SPAD Pixel Design Failure

A set of 'Deep SPAD' (deep N-well to P- Substrate) pixel trials could not be tested due to a design failure in the common front end circuit. As shown in Figure A.1.4, the DC bias of node V2 is critical to the circuit performance of this circuit which could not be maintained during the intended operating cycle. The pull-up was pulsed to raise the V2 DC level, unfortunately due to the leakage on that node the voltage would not remain constant during an exposure. During a SPAD pulse where the SPAD excess bias was greater than the DC level of the V2 node, the drain diodes of the NMOS devices (marked in red) would become forward biased for a short period and exacerbate the problem of node DC stability. Both the pull up and the disable pulse were locally re-generated on-chip and could not be isolated or tuned. These Deep SPAD circuits were not further explored.

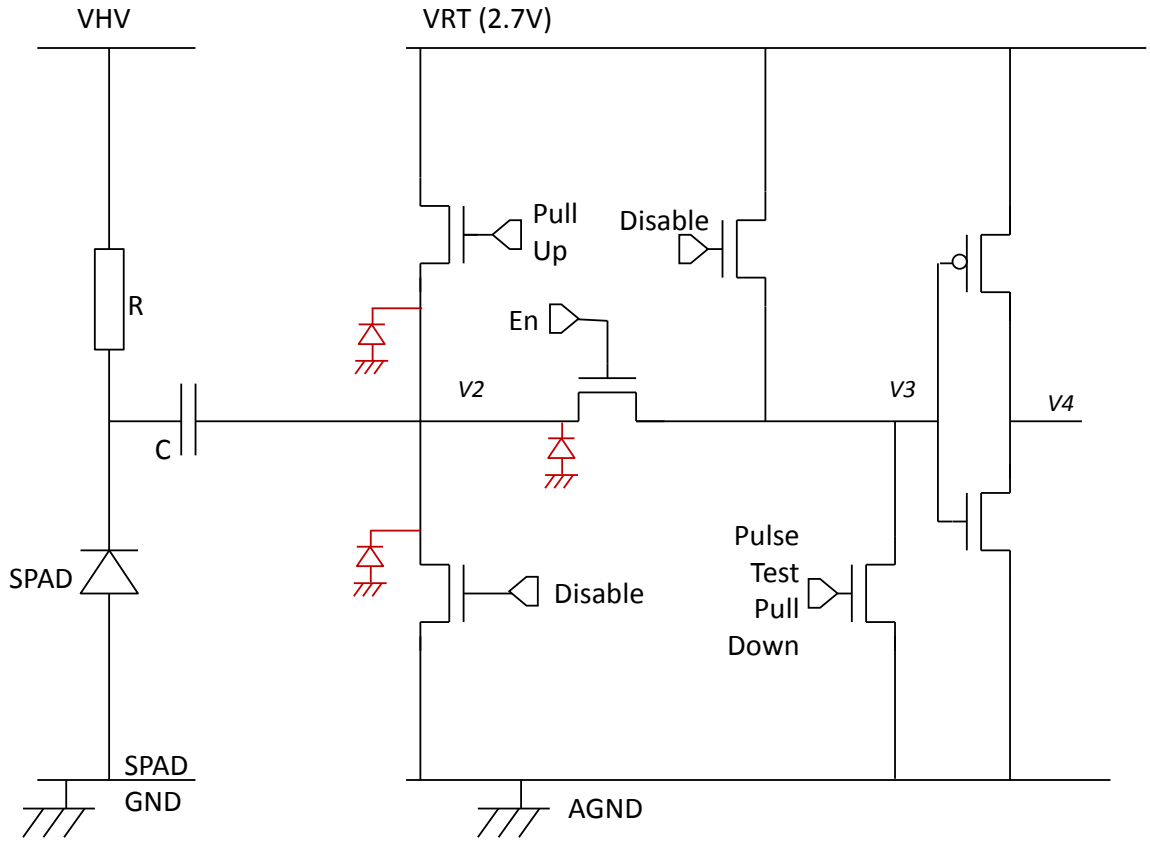


Figure A.1.4. Common Deep SPAD Front End Circuit with instability failure mode.

Appendix 2. TCSPC Sensor with 14GS/s Direct to Histogram Time to Digital Converter

This appendix gives a brief overview of the research project, undertaken in parallel with this doctoral thesis research, on multiple event TDC design for TCSPC applications such as optical ranging or LIDAR. The TDC architecture conceived replaces the conventional two step TCSPC process which is iterated for every recorded photon of time-code generation, followed by RAM value lookup, increment and write, into one parallel direct-to histogram process. An FPGA delay-line TDC with direct to histogram output was presented in [3] and is the highest throughput FPGA-based TDC in the known literature. An ASIC folded flash TDC implementation is published in [5], it obtains an order of magnitude increase in conversion rate over other ASIC-based TDCs and is described briefly in this appendix.

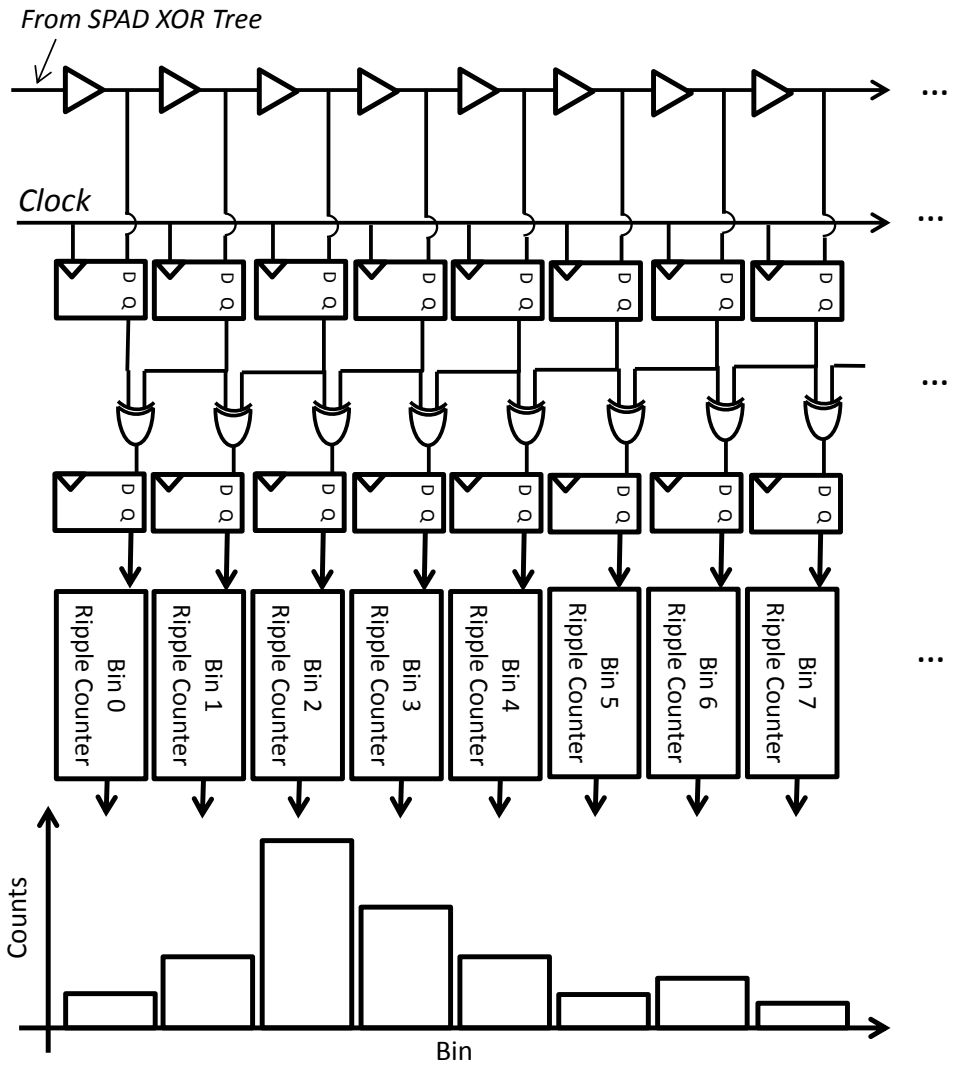


Figure A.2.1. The direct to histogram TDC concept applied to a reverse start stop delay-line flash TDC.

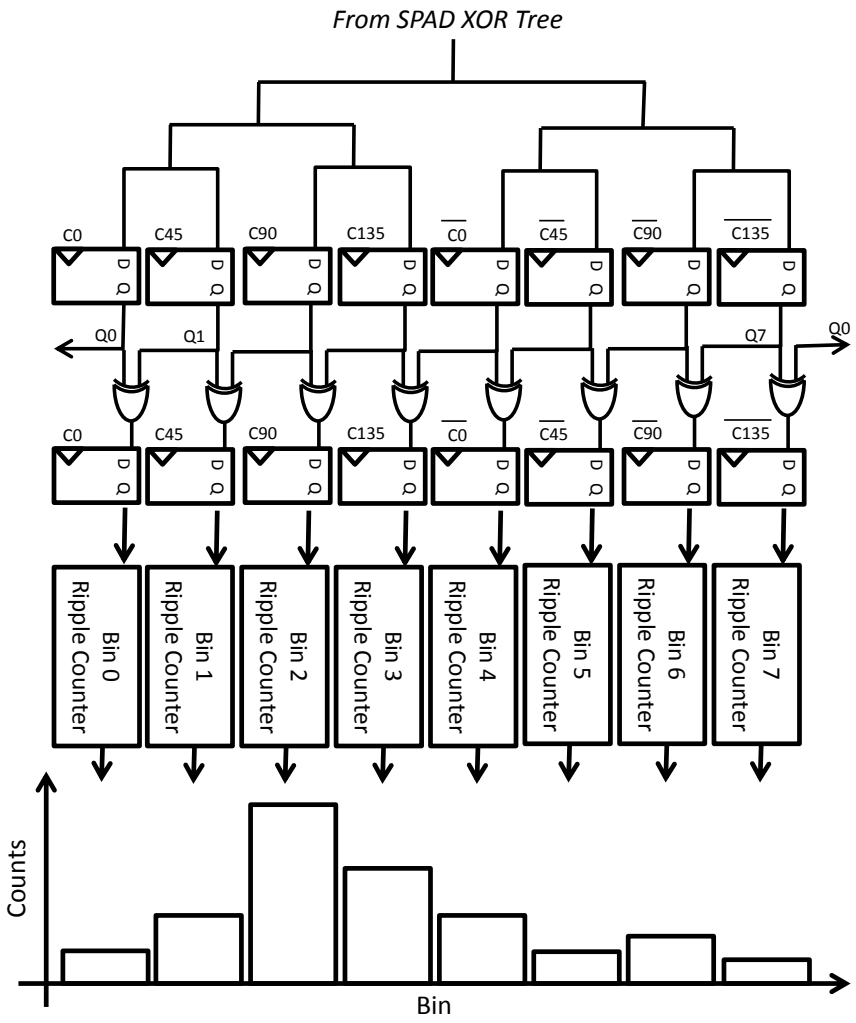


Figure A.2.2. The direct to histogram TDC concept applied using a forward-mode folded flash TDC example with an 8-phase PLL.

The direct histogram output TDC design is applied to the two structures of flash TDC: reverse-mode (stop to start) based on an input delay line with global clock, and forward-mode (start to stop) based on a global input and multiple-phase clocks from a PLL. Figure A.2.1 illustrates the reverse mode concept implemented in [3] parallelising the TDC outputs of a conventional delay-line TDC. The XORs provide transition detection outputting a multiple-hot thermometer code (a logical high from the XOR indicates the arrival of a photon at a particular time index). Counters sum up the number of transitions detected and the counter values represent a temporal histogram. Figure A.2.2 shows the same XOR logic and counters logic with a forward mode flash TDC front end based on an eight clock phase PLL as an example.

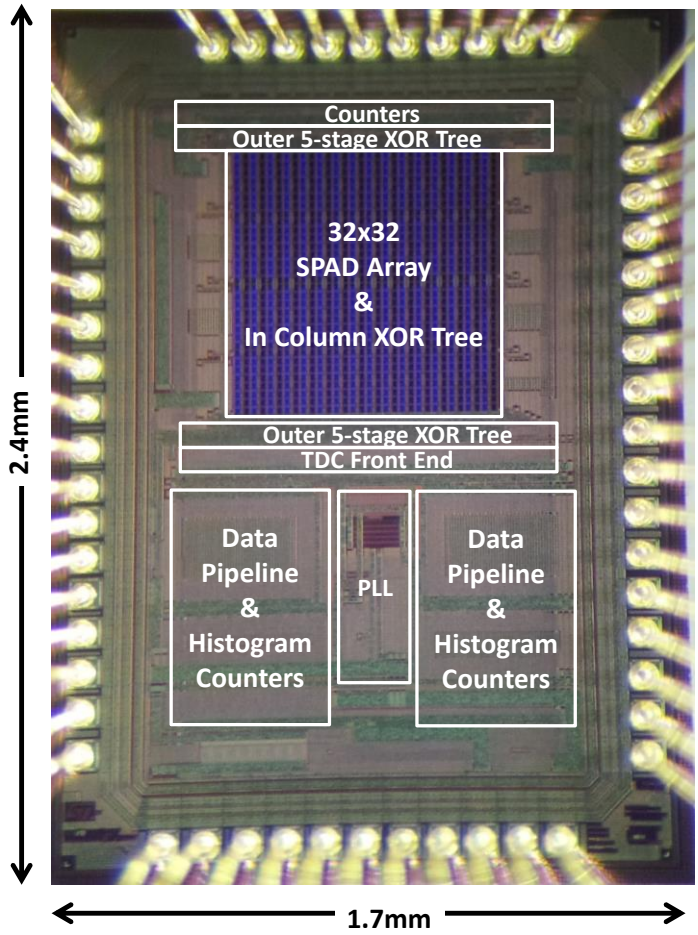


Figure A.2.3. Photomicrograph of the 14GS/s TCSPC sensor in a CPGA68 package with glass lid. The die measures 2.4mm x 1.7mm in ST's 130nm imaging 1P4M process.

The ASIC folded flash TDC published in [5] is shown in an annotated photomicrograph in figure A.2.3. The IC comprises a digital SiPM array of 32x32 'deep' SPADs (deep N-well to P- Substrate), a 1024 to 1 signal combining XOR H-tree, a 14GS/s direct to histogram TDC for TCSPC and a set of time-interleaved asynchronous counters for single photon counting. A block diagram of the single point TCSPC sensor is shown in figure A.2.4. The SPAD pixel consists of a 100kΩ poly resistor quench, decoupling capacitor (designed using the MOM capacitor design from chapter 3), input buffer, toggle flip-flop (TFF), and a 6T SRAM. The SRAM disables the input buffer and TFF but does not disable the SPAD. The pixel contains a 'deep' SPAD (described in [48]) that was designed to attain 43% fill factor in the 21µm pixel pitch. Each SPAD avalanche triggers the input buffer and flips the state of the TFF. This encodes the timing information of the SPAD leading edge on both positive and negative outputs from the pixel in an asynchronous dual data-rate (DDR) approach. These DDR pulses are combined through a ten-stage timing balanced XOR H-tree. The first five stage of the tree are single ended 8T XORs distributed in the array in a column-wise arrangement with centrally tapped output. The second final five stages are fully differential in a horizontal arrangement.

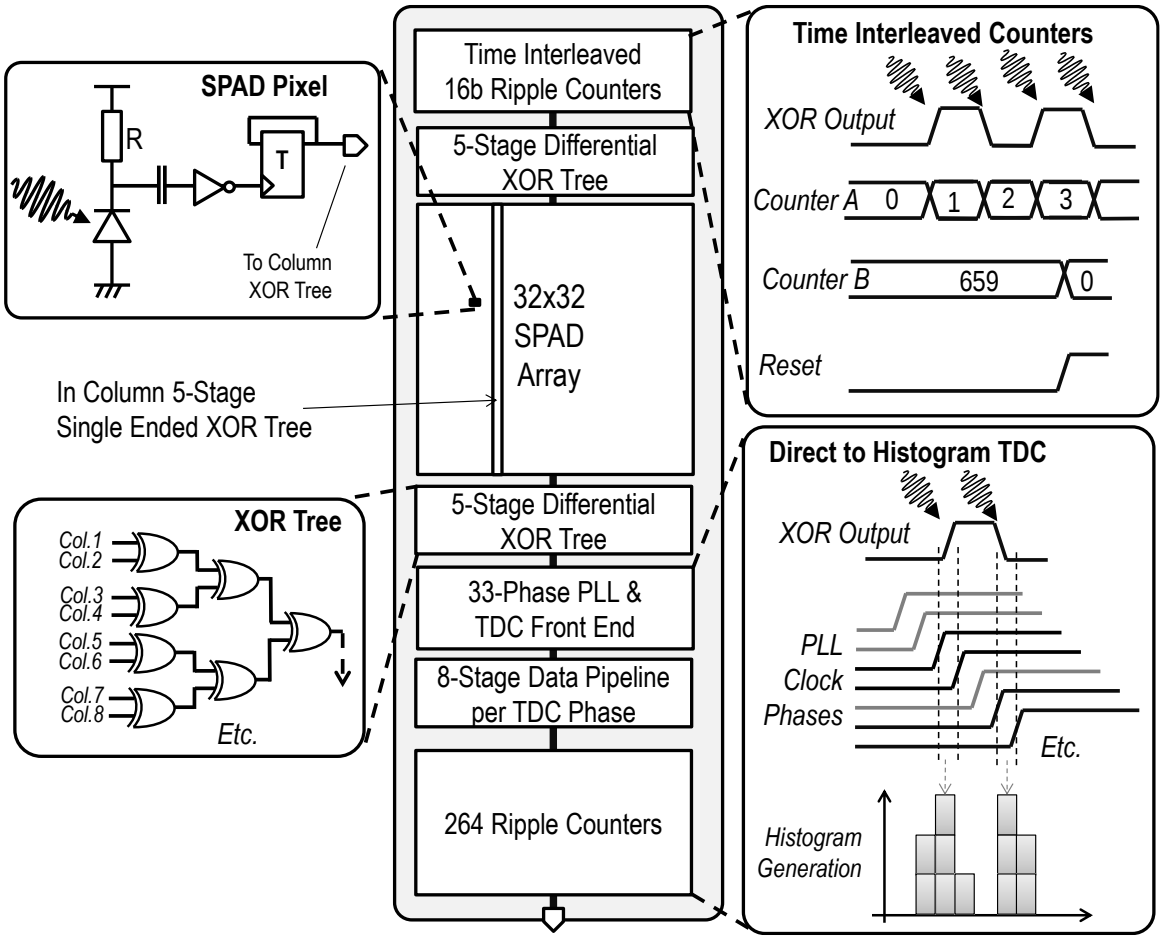


Figure A.2.4. Block diagram of the TCSPC sensor.

Figure A.2.5 shows a block diagram of the folded flash TDC. It consists of a PLL with 33-phase pseudo differential VCO. The VCO is designed to have an odd number of stages using only positive clock edges to protect the time conversion from duty-cycle variations. For each of the 33 clock phases, a flash TDC front end, XOR transition detection, eight stage shift register, and eight 16b ripple counters which directly build a histogram.

The difference between two successive positive clock edges creates a sampling window in time. As the final clock is folded back to the first this is a continuous operation with no conversion dead time or converter dead zone in the temporal dynamic range. The output of each front end flip-flop is XOR'd with the output from the next clock phase to detect a logical change in the SPAD array output representing a photon arrival. The throughput of the device is the reciprocal of the sampling window. The minimum sampling window or 'bin width' was measured at 71.4ps corresponding to 14GS/s conversion rate. An example histogram is shown in Figure A.2.6.

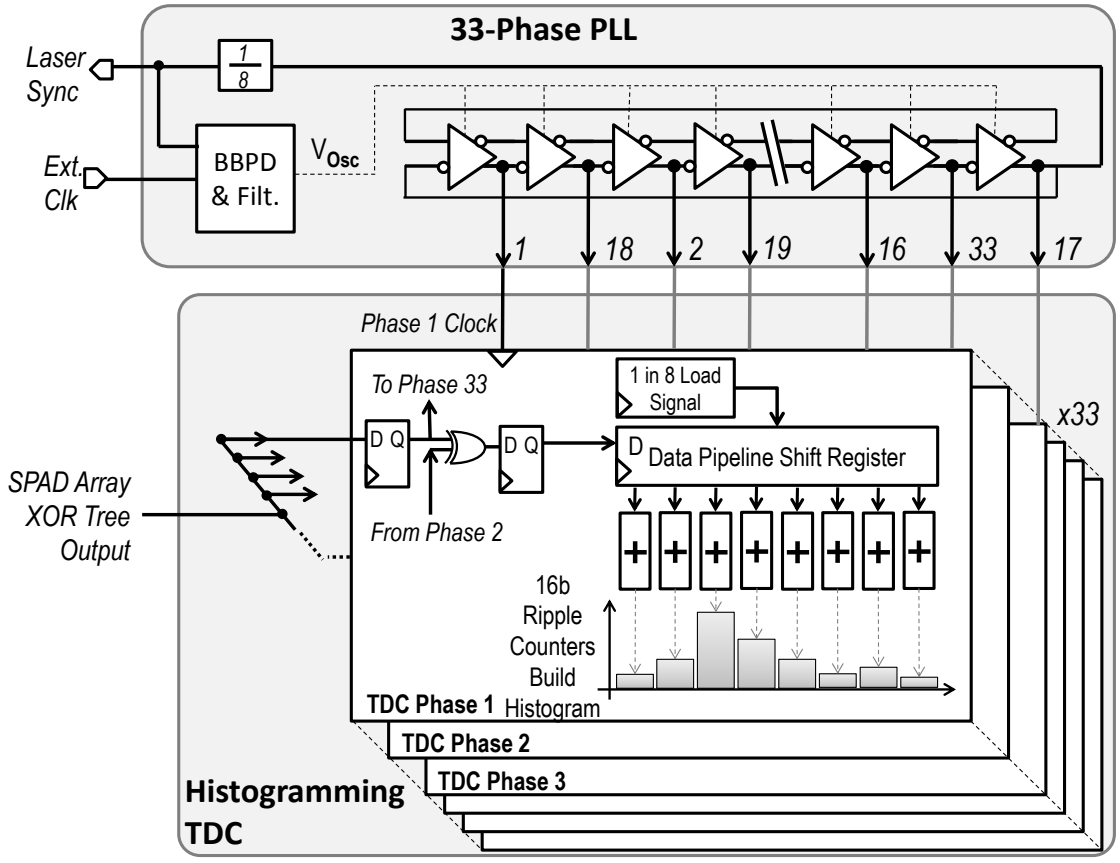


Figure A.2.5. Block diagram of the Folded Flash TDC with Direct to Histogram Output.

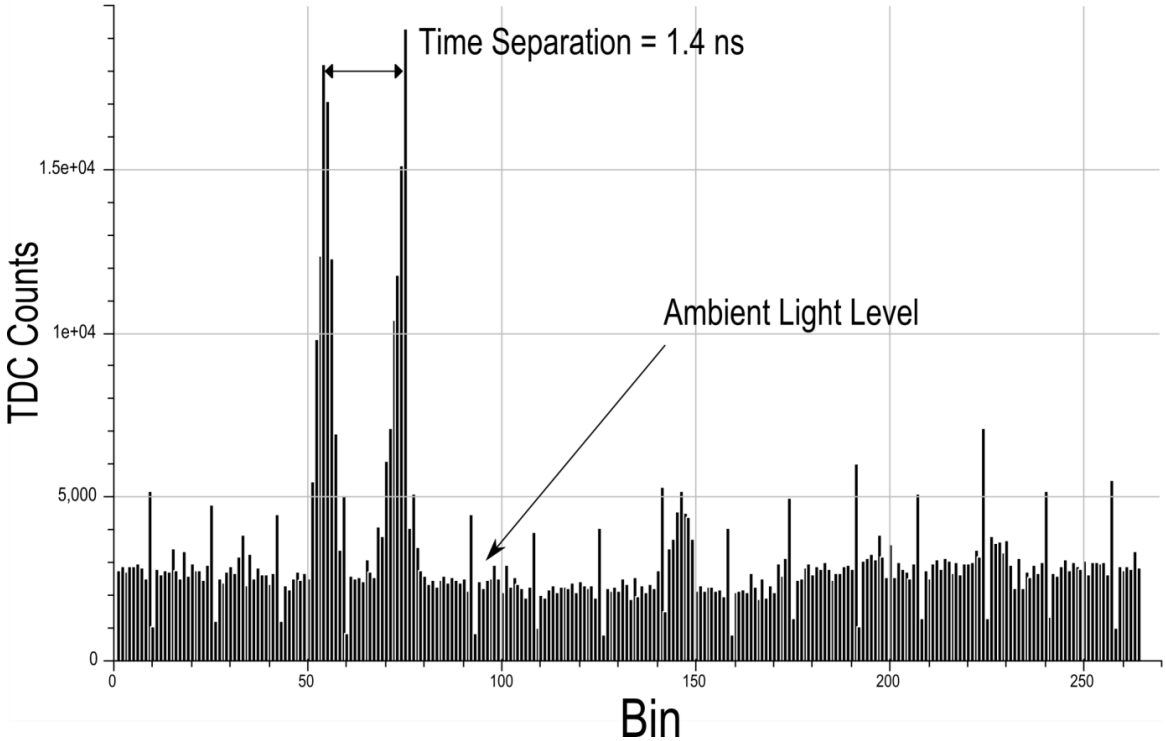


Figure A.2.6. Example histogram from the TCSPC sensor with two peaks from two lasers. Non-linearity spikes are evident in the ambient/background noise.

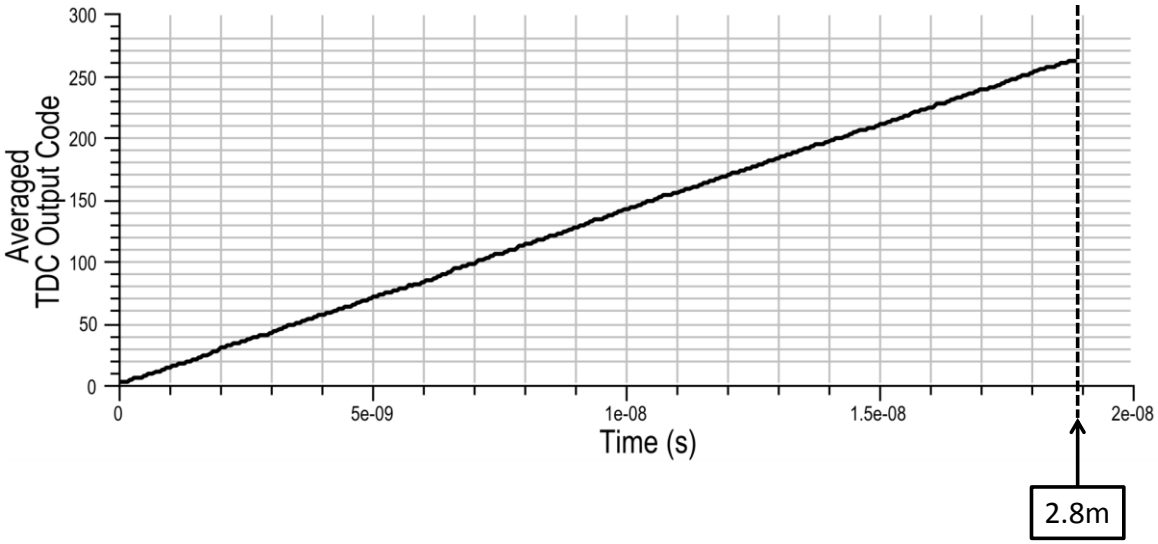


Figure A.2.7. Plot of averaged TDC code versus time showing a 2.8m LIDAR dynamic range at 71.4ps bin width with 264 histogram bins.

Figure A.2.7. shows the recorded average TDC code against time using a digital delay generator to incrementally delay a 425nm 80ps-FWHM PicoQuant laser across the 18.8ns dynamic range (equivalent to 2.8m optical ranging) showing no conversion dead zone.

Continuous single photon counting, using the asynchronous 16b ripple counters, is demonstrable using this IC for applications such as a fibre or free space optical receiver using visible light communication (VLC) or 'LiFi' and investigated in [9]. Figures A.2.8 and A.2.9 show the recorded photon counts from an experiment using a 425nm commercial LED, placed 5cm above the sensor, and connected to a sinusoidal signal generator for two different input frequencies (10kHz and 500kHz).

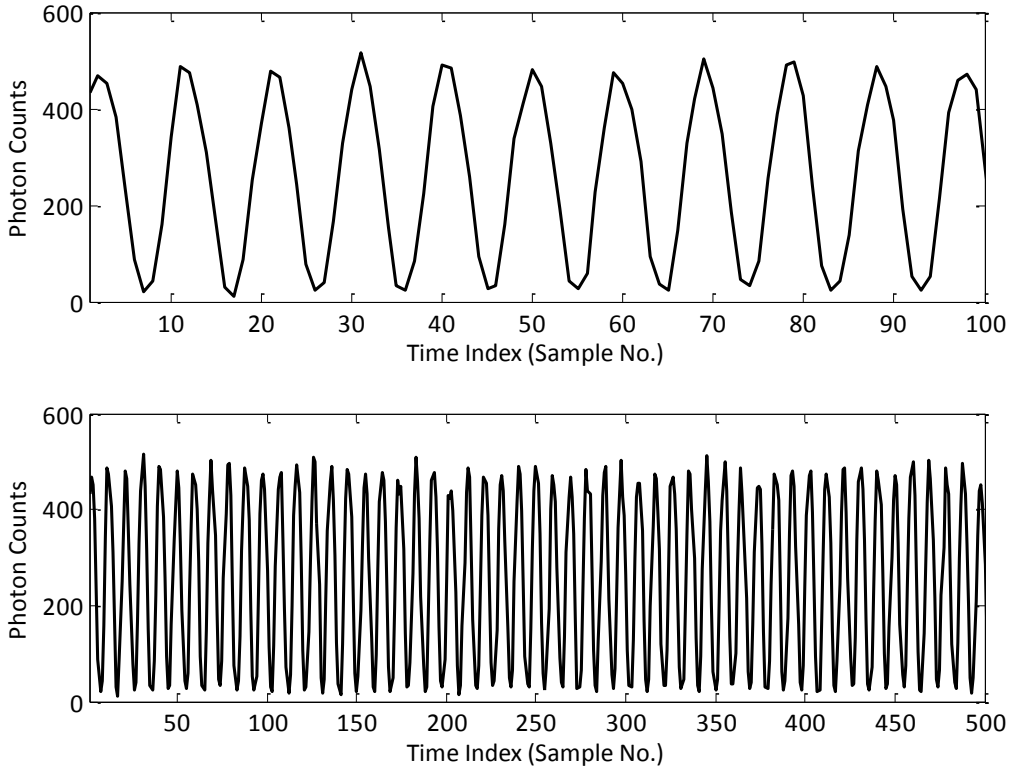


Figure A.2.8. Free space optical communications demonstration. A 10kHz signal input to an LED placed 5cm from the sensor. The sampling rate is 50kHz (reciprocal of 20μs counting window).

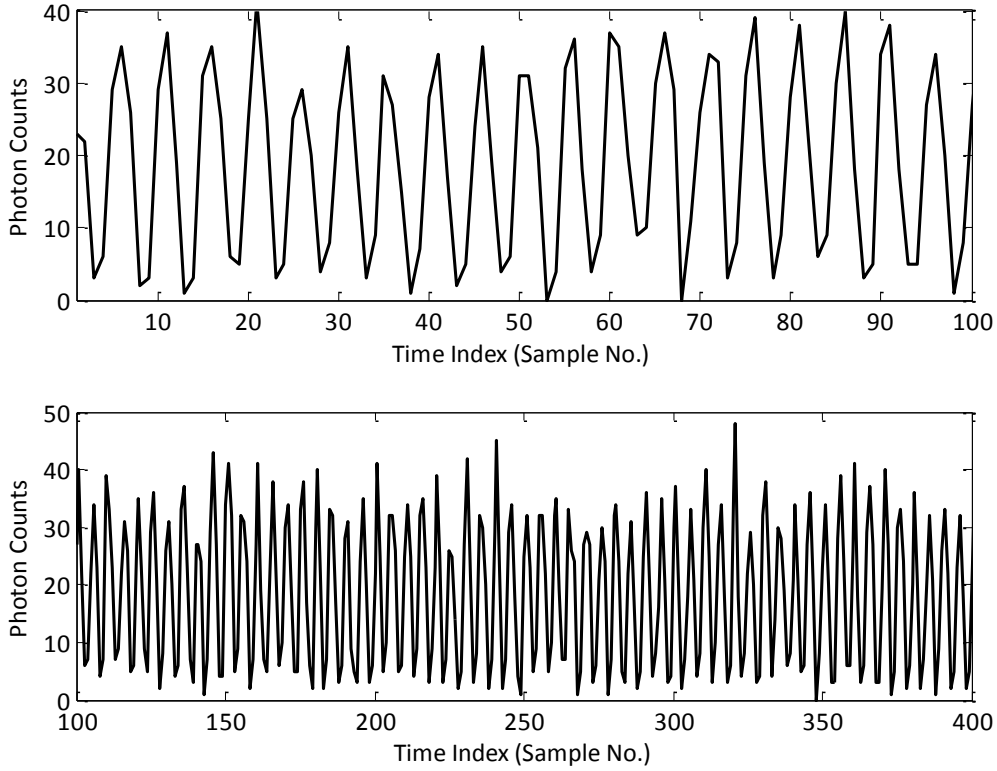


Figure A.2.9. 500kHz LED signal sampled at 5MHz (reciprocal of 200ns counting window).

Appendix 3. Sensor Characterisation and Development Platform

To support the testing and characterisation of this doctoral research and the work of colleagues in the research group, a characterisation platform was created to support the many test arrays, single point sensors and image sensors. The whole system from software through to the pad ring of the IC's was developed under the banner of the 'CSS Platform'. The PCBs were designed to be non-specific to a sensor with a set of regulators and high voltage generation in order to be as generic as possible. The padrings of a number of sensors conform to a particular pin out arrangement allowing connection to these common PCB's.

Mathwork's MATLAB software was chosen for the dual purposes of sensor interface, characterisation and data analysis. An interface was developed to ease communication, lower development time and improve human readability of the software interface to the prototype sensors. This consists of MATLAB code to wrap up calls to the Opal Kelly API and complementary automatically-created Verilog to wrap up the Opal Kelly Verilog instances. During the course of this research, Luca Parmesan and Aravind Venugopalan developed this further to create the matching Python-based control and wrapper code. A 'TCL' script was developed primarily by the author to create the MATLAB, Verilog and Python files. The script was written to handle both USB 2 and 3 computer interfaces. Figure A.3.1 presents a diagrammatic overview of the sensor characterisation platform.

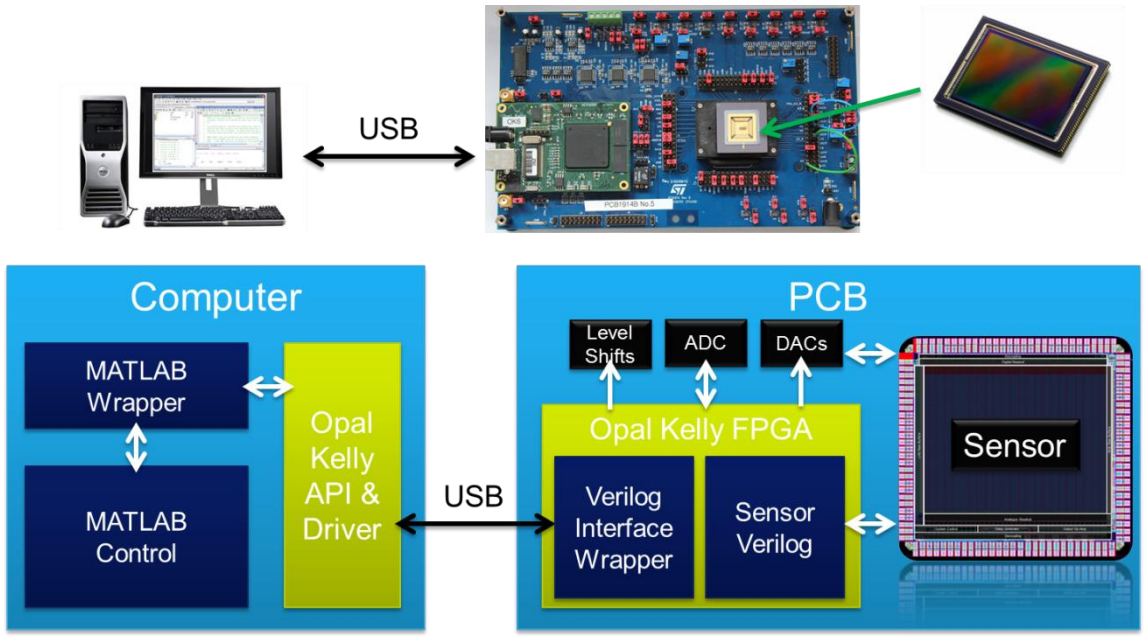


Figure A.3.1. Overview diagram of the 'CSS Platform' sensor characterisation and development platform.

A.3.1. Printed Circuit Boards

Three PCBs were created during the course of this research. The schematics of these are included in the later part of this appendix. The naming convention used (e.g. 'PCB 1914A') is from the ST Imaging division applications team who assisted in creating and in populating the components onto the boards. The PCB layout was done by an external PCB layout contractor 'CAD Energy'.

The three boards are listed with the respective sensors or test structures that were tested on them:

- PCB 1914 Revision A created in March to June 2012
    - 'TAC_PIXEL_TESTS' - Initial experimentation of the SPC test pixels described in Chapter 4 and [1].
- PCB1914 Revision B updated in May 2013.
    - 'TAC_PIXEL_TESTS' - TAC Test Pixels in [4] and described in Chapter 4.
    - 'PIXEL_CMM' - Integrating TAC and SPC in [150]
    - 'FLASHTDC' - Folded Flash Multiple Event TDC in Appendix 2.
- PCB1919 Revision A created in June 2013.
    - 'SPCIMAGER' - 320x240 SPAD-based SPC Image Sensor described in Chapters 5 and 6.
    - 'TACIMAGER' - 256x256 SPAD-based TCSPC Image Sensor described in [7].

Figure A.3.2 shows a photograph of PCB1914 revision A showing the two-connector socket on the left with white outline for the off-the-shelf Opal Kelly FPGA. At the top of the board from left to right, are the SPAD high voltage generation, three 20MS/s 14b differential ADCs, and regulators and DACs to supply the many power supplies and bias voltages and currents to the prototype sensors packaged in 68 pin ceramic pin grid array (CPGA) packages. Test jumpers for each of the CPGA pins surround the CPGA68 socket.

Figure A.3.3 shows a photograph of PCB1914 revision B containing a Xilinx Spartan 3 4000 series FPGA in green on the left. On the left side of the CPGA68 socket is a laser NIM-interfacing circuit and on the right is a fast BJT current mirror ramp generator circuit for TAC pixel characterisation.

Figure A.3.4 shows the image sensor development PCB 1919 revision 'A' mounted on a tripod.

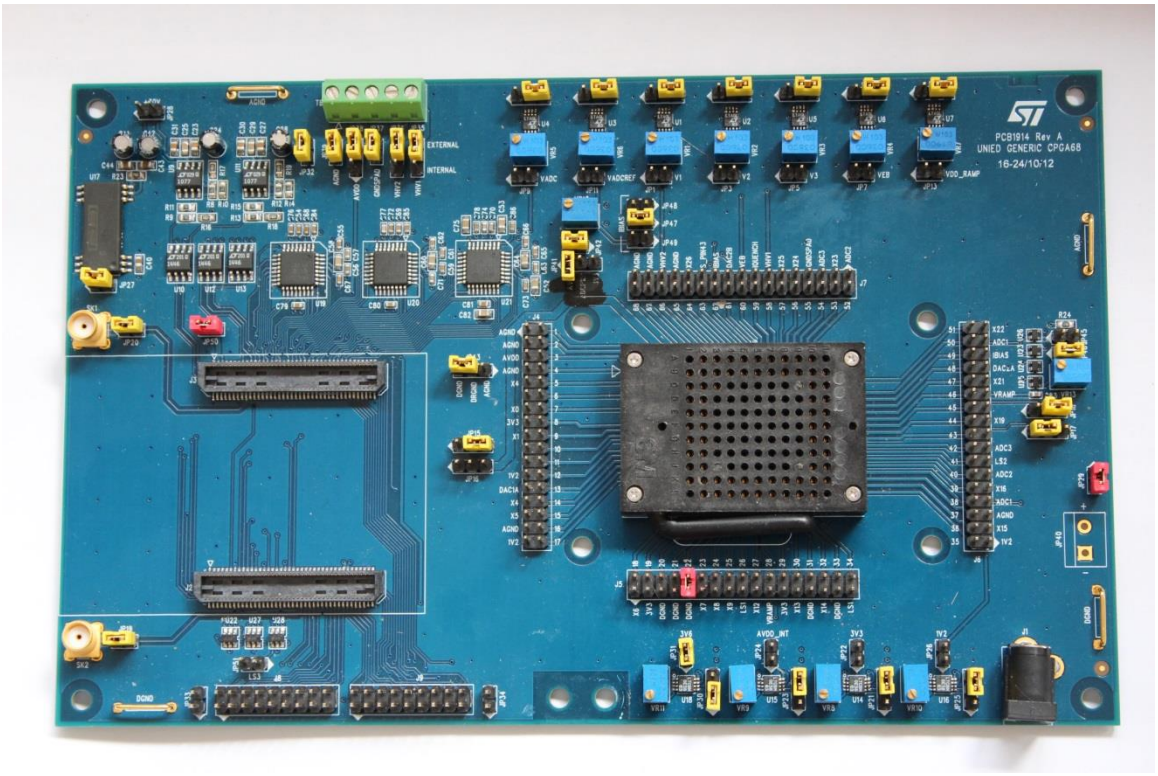


Figure A.3.2. Test array and TDC Characterisation PCB 1914 Revision A with CPGA68 Socket centre right and Opal Kelly socket on left.

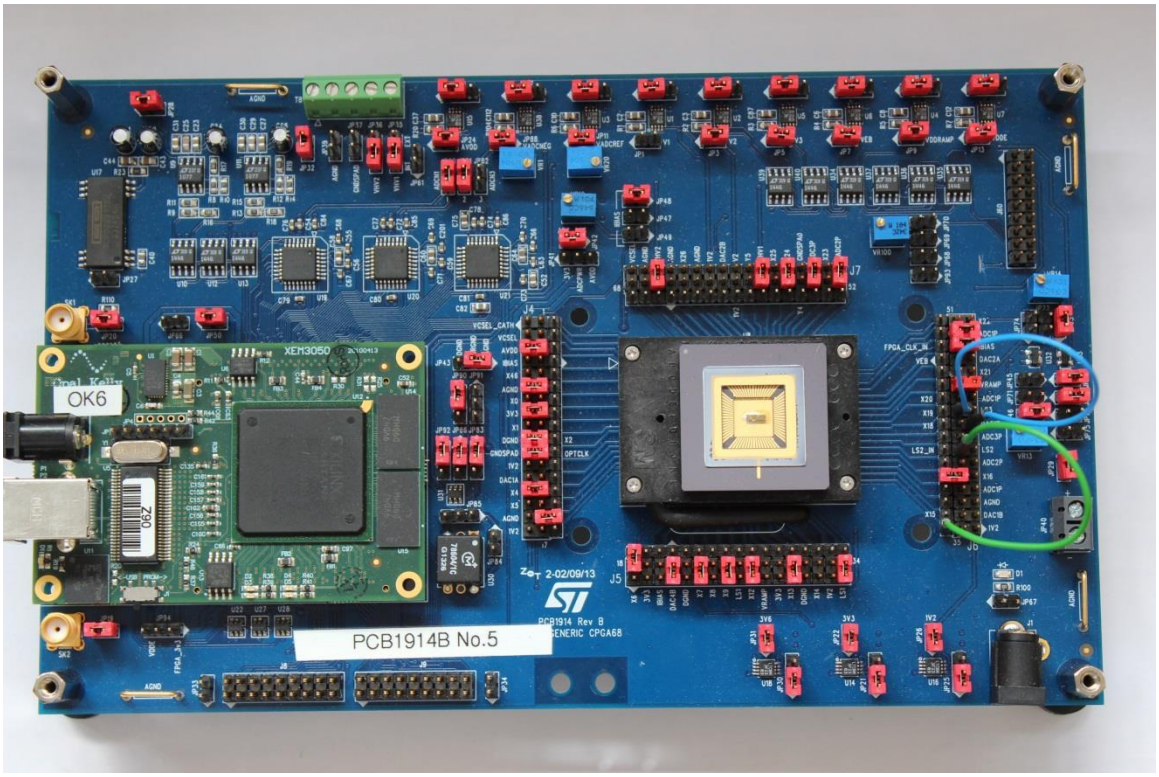


Figure A.3.3. Test array and TDC Characterisation PCB 1914 Revision B with test IC plugged into the CPGA68 Socket and Opal Kelly XEM3050 with Xilinx Spartan 3 4000 series FPGA mounted on the board.

Figure A.3.4. Image sensor characterisation PCB 1919 revision A with CPGA144 Socket and Opal Kelly XEM3050 with Xilinx Spartan 3 4000 series FPGA and STMicroelectronics custom made M12 lens holder.

A.3.2. PCB 1914 Revision A Schematics

The file for this schematic is included within the attached data disc and in the CSS Platform documentation folder under PCB. The file is located in the attached data disc:

- PCB Schematics / PCB1914A / PCB1914A.pdf

A.3.3. PCB 1914 Revision B Schematics

The file is located in the attached data disc:

- PCB Schematics / PCB1914B / PCB1914B.pdf

A.3.4. PCB 1919 Revision A Schematics

The file is located in the attached data disc:

- PCB Schematics / PCB1919A / PCB1919A.pdf

A.3.5. PCB 1919A CPGA144 Pin Out and SPCIMAGER Pad Ring Mapping

CPGA Side	CPGA Pin	Signal Category	PCB Signal	SPCIMAGER_AA & AB Pad	Description
Left	109	CORE POWER	1V2	vdd	1.2V Supply
	110	CORE POWER	AGND	gnd	
	111	CORE POWER	V1	VDD3V3	Quiet Analogue 3.3V
	112	CORE POWER	AGND	AVSS	
	113	CORE POWER	V2	VDD2V7	Pixel Supply VRT 2.7V
	114	CORE POWER	AGND	VSSPIXEL	
	115	DIGITAL	X0		
	116	DIGITAL	X1		
	117	DIGITAL	X2		
	118	RING POWER	VDDE	vdde3v3	Digital IO 3.3V
	119	RING POWER	DGND	gnde	
	120	DIGITAL	X3	RowAddr0	Row Address
	121	DIGITAL	X4	RowAddr1	
	122	DIGITAL	X5	RowAddr2	
	123	DIGITAL	X6	RowAddr3	
	124	DIGITAL	X7	RowAddr4	
	125	DIGITAL	X8	RowAddr5	
	126	DIGITAL	X9	RowAddr6	
	127	DIGITAL	X10	RowAddr7	
	128	RING POWER	VDDE	vdde3v3	Digital IO 3.3V
	129	RING POWER	DGND	gnde	
	130	DIGITAL	X11		
	131	DIGITAL	X12		
	132	DIGITAL	X13		
	133	DIGITAL	X14	ENAOUT	Time Gate Outputs
	134	DIGITAL	X15	ENBOUT	
	135	DIGITAL	X16		
	136	DIGITAL	X17		
	137	DIGITAL	X18		
	138	RING POWER	VDDE	vdde3v3	Digital IO 3.3V
	139	RING POWER	DGND	gnde	
	140	DIGITAL	X19	LatchAddress	Row /Col Address Latch
	141	DIGITAL	X20	ColAddr0	Col Address
	142	CORE POWER	AGND	AVSS	Quiet Analogue Ground
	143	CORE POWER	DGND	gnd	1.2V Supply
	144	CORE POWER	1V2	vdd	
Bottom	1	CORE POWER	V3	VDD3V6	Quiet Analogue 3.6V
	2	CORE POWER	V2		
	3	CORE POWER	V1		
	4	DIGITAL	X21		
	5	DIGITAL	X22		
	6	DIGITAL	X23	PIXRST	Pixel Reset
	7	DIGITAL	X24	DIS	Pixel Disable
	8	DIGITAL	X25	ColAddr1	Col Address
	9	DIGITAL	X26	ColAddr2	
	10	DIGITAL	X27	ColAddr3	
	11	DIGITAL	X28	ColAddr4	
	12	RING POWER	VDDE	vdde3v3	Digital IO 3.3V
	13	RING POWER	DGND	gnde	
	14	DIGITAL	X29	ColAddr5	Col Address
	15	DIGITAL	X30	ColAddr6	
	16	DIGITAL	X31	ColAddr7	
	17	DIGITAL	X32	ColAddr8	

CPGA Side	CPGA Pin	Signal Category	PCB Signal	SPCIMAGER_AA & AB Pad	Description
Bottom	18	DIGITAL	X33	BINAENABLE	Time Gate Input
	19	DIGITAL	X34	BINBENABLE	
	20	ANA	OPTCLK	OPTCLK	Laser Sync Input
	21	DIGITAL	X35		
	22	SPAD HV	VHV	VHV	SPAD High Voltage Supply
	23	SPAD HV	GNDSPAD	VSSSPAD	
	24	SPAD HV	VHV	VHV	
	25	SPAD HV	GNDSPAD	VSSSPAD	
	26	SPAD HV	VHV	VHV	
	27	SPAD HV	GNDSPAD	VSSSPAD	
	28	SPAD HV	VHV	VHV	
	29	SPAD HV	GNDSPAD	VSSSPAD	
	30	SPAD HV	VHV	VHV	
	31	SPAD HV	GNDSPAD	VSSSPAD	
	32	RING POWER	VDDE	vdde3v3	Digital IO 3.3V
	33	RING POWER	DGND	gnde	
	34	DIGITAL	X37	GS	Global Shutter (GS) Mode
	35	DIGITAL	X38		
	36	DIGITAL	X39		
Right	37	DIGITAL	X40	GS_BIN_TOGGLE	GS Bin Toggle Select
	38	ANA	DAC6		
	39	ADC	ADC1_P	ANA_OUT_N	ADC Output
	40	ADC	ADC1_N	ANA_OUT_P	
	41	ADC	ADC2_P		
	42	ADC	ADC2_N		
	43	CORE POWER	V4	VDDOPAMP	Op Amp Power Supply
	44	CORE POWER	AGND	VSSOPAMP	
	45	DIGITAL	X41		
	46	DIGITAL	X42		
	47	ANA	IBIAS1	I_OPAMP_BIAS	Bias Currents
	48	ANA	IBIAS2	I_COL_BIAS	
	49	CORE POWER	V5		
	50	CORE POWER	AGND	VSSPIXEL	Pixel Ground
	51	ANA	DAC1	VS	Pixel Bias Voltages
	52	ANA	DAC2	VG	
	53	ANA	DAC3	VQ	
	54	CORE POWER	AGND	AVSS	Quiet Analogue Ground
	55	DIGITAL	X43		
	56	DIGITAL	X44	CROWBAR	Column Timing
	57	DIGITAL	X45	CDSBLK	
	58	DIGITAL	X46	CDSSIG	
	59	DIGITAL	X47	SR_CLK	Control Serial Interface
	60	DIGITAL	X48	SR_IN	
	61	DIGITAL	X49	SR_LOAD	
	62	RING POWER	VDDE	vdde3v3	Digital IO 3.3V
	63	RING POWER	DGND	gnde	
	64	DIGITAL	X50	RSTN	Reset_N
	65	DIGITAL	X51	ADCCLK	Digital Comparator Latch
	66	DIGITAL	X52	SERIAL	Digital Serialiser Load
	67	DIGITAL	X53	SERCLK	Digital Serialiser Clock
	68	ANA	DAC5		
	69	CORE POWER	V5		
	70	CORE POWER	AGND	VSSPIXEL	Pixel Ground
	71	CORE POWER	V2	VDD2V7	Pixel Supply VRT 2.7V
	72	CORE POWER	V2	VDD2V7	

CPGA Side	CPGA Pin	Signal Category	PCB Signal	SPCIMAGER_AA & AB Pad	Description
Top	73	CORE POWER	AGND	gnd	1.2V Supply
	74	CORE POWER	1V2	vdd	
	75	CORE POWER	V1	VDD3V3	Quiet Analogue 3.3V
	76	CORE POWER	AGND	AVSS	
	77	ANA	DAC4	VREF	Digital Comparator Reference
	78	DIGITAL	X54		
	79	DIGITAL	X55	DOUT0	Digital Serialiser Outputs
	80	DIGITAL	X56	DOUT1	
	81	DIGITAL	X57	DOUT2	
	82	DIGITAL	Y3	DOUT3	
	83	DIGITAL	Y4	DOUT4	
	84	DIGITAL	Y5		
	85	RING POWER	VDDE	vdde3v3	Digital IO 3.3V
	86	RING POWER	DGND	gnde	
	87	DIGITAL	Y6	DOUT5	Digital Serialiser Outputs
	88	DIGITAL	Y7	DOUT6	
	89	DIGITAL	Y8	DOUT7	
	90	DIGITAL	Y9	DOUT8	
	91	DIGITAL	Y10	DOUT9	Digital Serialiser Outputs
	92	DIGITAL	Y11	DOUT10	
	93	DIGITAL	Y12	DOUT11	
	94	DIGITAL	Y13	DOUT12	
	95	RING POWER	VDDE	vdde3v3	Digital IO 3.3V
	96	RING POWER	DGND	gnde	
	97	DIGITAL	Y14		
	98	DIGITAL	Y15		
	99	DIGITAL	Y16	DOUT13	Digital Serialiser Output
	100	DIGITAL	Y17	DOUT14	
	101	DIGITAL	Y18	DOUT15	
	102	DIGITAL	Y19		
	103	SPAD HV	VHV2	VHV2	Test SPADs High Voltage Supply
	104	SPAD HV	VHV2	VHV2	
	105	SPAD HV	GNDSPAD	VSSSPAD	
	106	SPAD HV	GNDSPAD	VSSSPAD	
	107	CORE POWER	V3	VDD3V6	Quiet Analogue 3.6V

Appendix 4. Videos

This section details the contents of the video files included within the attached data disc. The following videos have been captured using the SPC Imager 'AB' in the oversampled binary QIS mode by Dr Istvan Gyongy:

- balloon_raw_bit_planes.avi – 1 bit binary, raw data from the sensor of a balloon bursting.
- balloon_fix_sum_4.avi – 2bit fixed sampling of a balloon bursting.
- balloon_fix_sum_16.avi – 4bit (16 bit planes averaged) fixed sampling of a balloon bursting.
- beads_raw_bit_planes.avi – 1 bit binary, raw data from the sensor of polystyrene beads.
- beads_fix_sum_4.avi – 2bit fixed sampling of falling polystyrene beads.
- beads_fix_sum_16.avi – 4bit fixed sampling of falling polystyrene beads.
- beads_rol_vs_fix_sum_16.avi – 4bit rolling versus fixed oversampling of beads.
- fan_raw_bit_planes.avi – 1 bit binary, raw data from the sensor of a rotating fan.
- fan_fixed_vs_rolling_32_sum.avi – 5bit rolling versus fixed oversampling of a rotating fan.
- fan_fixed_vs_rolling_128_sum.avi – 7bit rolling versus fixed oversampling of a rotating fan.
- fan_adaptive_128_32_sum.avi - 7bit rolling versus adaptive rolling 5bit/7bit oversampling.
- milk_raw_bit_planes.avi – 1 bit binary, raw data from the sensor of milk spilling.
- milk_fix_sum_4.avi – 2bit fixed sampling of milk spilling.
- milk_fix_sum_16.avi – 4bit fixed sampling of milk spilling.

Appendix 5. 3D Vision Research Background

There are a number of optical techniques for extracting distance information from a scene. There are two distinct approaches: 'active' which uses an illuminator as a reference (either temporally or spatially) and 'passive' which depends on a constant and uniform illumination on a scene often depending solely on ambient light. Active techniques consist of inferometry, structured light, divergence ratio ($1/D^2$) and time of flight (TOF). Passive techniques rely on the contrast of edge detection; these are stereoscopic, plenoptic and focus depth. In this appendix, these active and passive techniques are discussed for 3D vision.

A.5.1. 3D Vision

The human visual system is a remarkable confluence of optics, sensing, real-time image processing, and system adaption. By using binocular vision (from the Latin 'bini' meaning two together and 'oculus' meaning eye), the human brain takes the images from the left and right eyes and through this stereo perception, or stereopsis, it interprets horizontal image disparity in a scene in order for us to navigate and interact with the world around. This neural processing occurs quickly, allowing humans to make accurate and rapid distance measurements that facilitate visual guidance of movement. This form of stereoscopic analysis takes place in the horizontal meridian but not, of course, the vertical meridian.

During approximately the last twenty years, there has been much research and development into 3D vision systems to allow machines to do the same, to detect the world around them, by having depth perception comparable to the capability of humans and animals [62]. There are many examples of 3D vision sensors providing 'sight' to guide action to computer systems for automotive (both driver and road facing), 3D scanning, security, robotics (both consumer and industrial) and medical applications. Yet, it is worth noting that we are still at the beginning of 3D vision research; these twenty to thirty years of development are dwarfed by the capacity of the human visual system which has undergone millions of years of evolutionary 'research and development'.

A.5.2. Human Computer Interface

3D vision sensors are designed to serve two purposes which will be discussed in this and the following section: as a human computer interface (HCI) and for machine vision. Since the invention of the first commercially successful type-writer in 1881 [177], the human machine interface has been predominantly the keyboard. Nearly ninety years later in 1970, Douglas Engelbart invented the computer mouse allowing a more natural interaction between the hand and the computer screen [178]. Over the last thirty years there has been a rise of computing devices specifically for playing games. The interface for gaming has been primarily a two handed

controller – arguably an ergonomic and pared down redesign of the keyboard and mouse. As general computing and gaming have shifted to mobile phones and tablets, the physical keyboard, mouse and game controller have been replaced in these small devices by a person's finger touching the screen of the computing device. In desktops and laptops, this touch screen interface supplements the traditional keyboard and mouse. The most pervasive example of this technology is the Apple iPhone, which launched in 2007 having only a small number of buttons, the system operation relied upon the touch screen capability in stark contrast to the other smart phones on the market [179]. Apple created a common code for multiple finger interactions or 'multi-touch' gestures across their whole range of computing and mobile devices. The success of which is apparent, as their suite of common gestures has been widely adopted by other manufacturers such as pinching in or out to zoom, or swiping to move sideways [180]. Yet in this current paradigm, if a user lifts their hand away from the keyboard, mouse, game controller or touch screen then the interaction stops. By using 3D vision sensors to augment the computer interface, this allows the interaction to continue (ideally with seamless transition) as a user becomes physically distant from the device. Pushing this concept further, for those computing applications that involve no physical interaction, the 3D vision sensor becomes the sole interface. Microsoft launched their 'Kinect' 3D vision sensor (designed and manufactured by PrimeSense, now owned by Apple [181]) to accompany their 'Xbox 360' gaming console in 2010 announcing a new concept of game play without the physical game controller. Under the banner 'You Are the Controller', consumers were encouraged to use whole body gestures to control the game [182]. The addition of a 3D vision sensor with a games console is now a competitive feature as seen in both the latest incarnations of the Microsoft Xbox and Sony PlayStation product lines [183], [184].

However, for greater and wider uptake of this gesture recognition technology, a hurdle must be overcome; unlike touch-screens, there is not yet a standard of human gestures for controlling devices. Like the rise and then dominance of Apple's mobile products establishing a multi-touch standard, it will take a major player in the consumer market to champion a gesture recognition standard. Alternatively and just as likely, an international agreement may be formed through the ISO or IEEE standards bodies.

The physical size of many current 3D vision sensors on the market prevents immediate transition into mobile phones or tablets. Recent reductions in size have recently occurred in 3D vision sensors notably in the SoftKinetic DS536 and PMDTec CamBoard Pico XS [185], [186]. Yet these are still physically too large to be integrated in mobile devices as front facing devices (on the same side as the touch screen). These sensors will need to be decreased in size by an order of magnitude and the number of components reduced to fit into a mobile phone or a tablet device.

If the hurdles of gesture language standardisation, size, power and importantly consumer demand are crossed, then 3D vision sensors have potential to be globally adopted as part of standard human computer interaction.

A.5.3 Requirements for 3D Vision Sensors

Regardless of the differing sensor technologies, there is a common set of needs for any 3D vision system. Figure A.5.1 shows an outline list of these needs in terms of the optical sensor, the system integration (combining the sensor, optics, and if required the illuminator) and the output data (in the form of a depth map or point cloud).

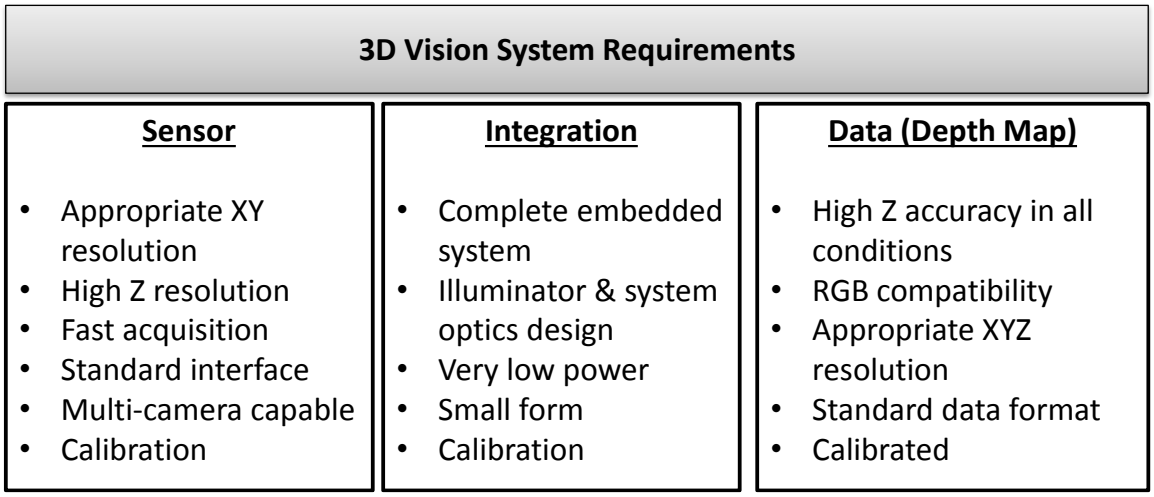


Figure A.5.1. Requirements for 3D vision sensors, system integration and output data.

Whether a simple proximity sensor or body tracking 3D depth camera, all vision systems must be able to detect a target in any environmental condition. These external conditions can be grouped into three sets:

- Ambient: the level of background light and the rate of change of the background light. Other external or environmental signals or noise sources such as rain, snow, fog.
- Target: the target reflectance and the distance separation between target and background.
- Movement: the absolute movement of the sensor and the relative movement of the target.

3D vision sensors are split into two types: those that use an active illumination device (in the most basic form, a continually operating LED) and those that have no illumination and passively rely on external light sources. The external condition scenarios which can be both most and least difficult for the sensor to resolve, are different between active and passive devices. Figures A.5.2 illustrates the worst and best cases for active illumination systems for the three groups of external conditions and indicates the needs of the system in order to handle these. Figure A.5.3 shows the comparative table for passive systems.

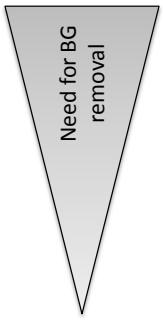
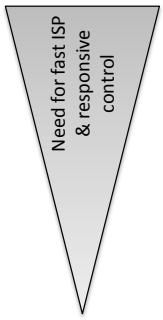
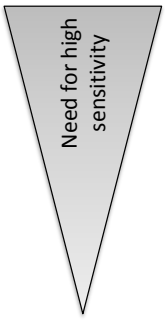
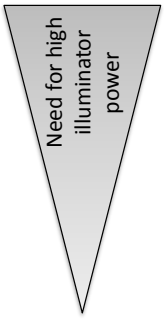
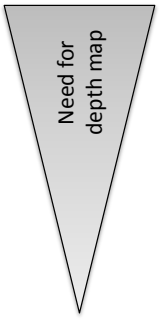
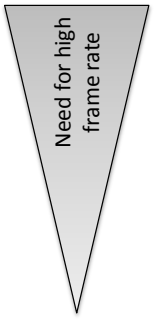
	Ambient		Target		Movement	
	Level	**Variation**	**Type / Reflectance**	**Distance**	**Absolute Sensor**	**Relative Target**
Worst Case	High (Sunny)	Fast	Dark /Low	Far	Moving	Moving
System Need						
Best Case	Low (Dark / Indoor)	Slow	Mirrored or White /High	Near	Stationary	Stationary

Figure A.5.2. Active Illumination: Worst to best case scenarios for 3D vision sensors using active illumination systems showing the degree of need for different systems requirements.

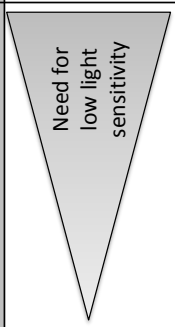
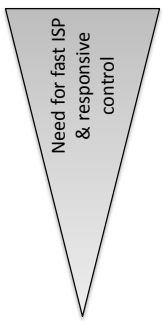
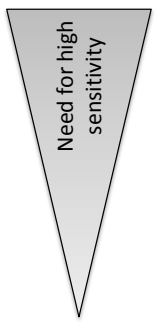
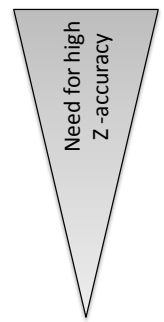
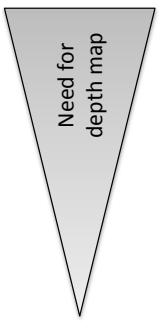
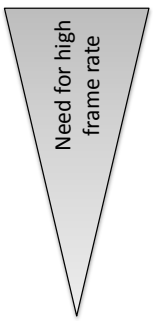
	Ambient		Target		Movement	
	Level	**Variation**	**Type / Reflectance**	**Separation (target to background)**	**Absolute Sensor**	**Relative Target**
Worst Case	Low (Dark/Indoor)	Fast	Dark /Low	Small Separation	Moving	Moving
System Need						
Best Case	High (Sunny)	Slow	Mirrored or White /High	Well Separated	Stationary	Stationary

Figure A.5.3. Passive Illumination: The comparable table for 3D vision sensor scenarios for passive systems showing the degree of need for different systems requirements.

The correct choice of sensor technology and matched system level processing to meet these external challenges is paramount. There are two well-publicised examples of potentially costly and certainly embarrassing system performance when the wrong choice was made.

In developing a camera for a desktop computer for HCI and face detection, from a CIS technologist's point of view, a conventional CIS at a fixed focus with an edge detection ISP algorithm would appear to be the simplest solution as the technology is mature and it is easy to implement. However, this passive system fails a simple test; the system must be able to clearly distinguish a moving person with dark skin against a dark background. A face detection algorithm within a Hewlett Packard webcam (using the factory default settings) was able to detect and track white faces but not black faces in a well-lit static scene leading to a viral YouTube video and negative press coverage [187], [188].

The second example demonstrates the importance of accurate gesture recognition and human feature detection. The Nest smart fire and smoke alarm is fitted with a machine vision sensor to allow silencing of the alarm with a wave of a hand [189]. However, the company suspended sales and disabled this feature when it was discovered that any movement around the alarm (and not the intended hand wave) could deactivate the system – a potentially harmful or deadly system bug [190].

These two examples, showing systematic inability to handle different target and movement conditions, highlight the importance of rigorous testing and choice of the correct detector technology.

A.5.4. 3D Vision Sensor Technology

There are a number of technological approaches for 3D vision sensors that are well described in this extensive textbook on the subject [62]. The different methodologies are briefly discussed in this section evaluating their suitability in different applications. There are two groups of 3D vision technology: those devices based on the TOF principle and those based on regular CIS pixels.

A.5.4.1. TOF Cameras

This first group of TOF sensors are covered in detail in chapters two and three, and Appendix 6, and so are only briefly mentioned here for comparison with CIS-based system. These TOF sensors are based on:

Photo modulation device or photo mixing device (PMD): Based on a fast PIN photo-diode or a CCD-style photo-gate. Covered in Appendix 6.

SPAD or APD: Avalanche photo-diode (APD) or Geiger mode device (G-APD or SPAD). Covered in Chapters 2 and 3.

The second group based on regular CIS technology is not central to this thesis and so is only covered in this background appendix section. It consists of five main approaches.

A.5.4.2. Low Resolution CIS

A small array of CIS pixels forming a low resolution imaging device can be used for close distance applications such as proximity detection and basic gesture recognition. There are three types of proximity sensors on the market: intensity, divergence ratio and triangulation. Intensity based proximity detection sensors perform a rudimentary light to dark detection test and the output is single bit binary but provides no distance information [191]. The illuminator is a continual operation LED at the wavelength of interest. In divergence ratio devices, distance information is calculated from the inverse square law drop, or divergence ratio of the intensity of a target to a calibrated reflectance. However, the reflectance of a target would need to be known *a priori* otherwise a low reflectance target would register at a different distance to a high reflectance target if placed co-incident at the same distance from the sensor [192]. Furthermore, ambient light level would skew the distance result. Alternatively in a third method, distance information for a single point can be gained through triangulation using the parallax shift along a line sensor [191]. Furthermore, basic gesture recognition could be achieved with a small CIS pixel array with a lens and basic edge detection and edge tracking similar to a mouse sensor. However, this approach fails with dark targets or high ambient light and cannot cope with absolute sensor movement as the moving background may be interpreted as a gesture.

A.5.4.3. Stereo Cameras

Two high resolution CISs are placed at known physical separation (referred to as the sensor 'baseline') and capture the same scene. The spatial differences between object edges between the two images are used to triangulate their respective positions. The spatial resolution of the system depends on the camera separation and the clarity of the object edge. A scene with low contrast edges would render the system unable to resolve a depth map. An illuminator in this system is used as a camera flash to cope with low light scenarios. The downside to this system is the computation required per frame. Examples are seen in the Sony PlayStation 4 [184] , the stereo camera from New Imaging Technologies [193] and in the Leap Motion computer gesture controller [194].

A.5.4.4. Structured Light

A pattern of NIR dots is projected onto a scene and then imaged by a CIS. The deformation of dots is then computed to produce a depth map. This technology has been the basis one of the most prevalent 3D image

sensors on the market - the original Microsoft/Primesense 'Kinect' [195]. It is a computationally heavy procedure and so is not suitable for very high frame rate applications. Furthermore, as it uses a projector there is a classic engineering trade-off between physical size and the illuminator power which can be achieved in the physical package. The projector intensity is important as ambient or background light rejection worsens with lower illuminator power. When the ambient light is brighter than the projected pattern, no depth map can be created. This is unlike TOF cameras where the ambient light can be much brighter than the illuminator.

A.5.4.5. Plenoptic or Light Field Cameras

Plenoptic cameras are similar to the stereo approach but they use a single sensor and more than two perspectives. For the interested reader, Dr Ren Ng's (founder of Lytro) PhD thesis is a comprehensive text on the theory of light field cameras [196]. An array of microlenses (much larger than current wafer scale microlenses) is placed above the sensor and is used to create a number of sub-images across the sensor array forming a lenticular array like the eye of an insect [197]. The sub-images are then processed by means of a weighted data mask and horizontal and vertical shifting to match features between sub-images. This computation has two outputs, an image that can be focused at any or every distance (and can easily be re-focussed for more than one distance range) and calculates, as part of the ISP algorithm, the distance of objects from the image sensor through trigonometry and so creates a depth map [198]. The computation is the downside to this technology, requiring a GPU to handle the very large dataset created per frame. Only recently have light field camera manufacturers Raytrix and Pelican released products working at 30FPS [199], [200]. This is a promising technology for mobile imaging, as it allows re-focussing, depth maps and, most significantly, it reduces the camera module height.

A.5.4.6. Swept Focus - Phase or Contrast Detection

Discrete auto-focus (AF) line sensor IC's found in digital single lens reflex (DSLR) cameras use a passive phase detection method imaging an object through two separator lenses onto a line sensor [201]. The focus distance is found using the line sensor, sweeping the focus and stopping when the two imaged spots match a defined horizontal separation on the sensor plane. If the two images are below the horizontal distance, the focus of the DSLR lens is swept back toward the camera. Conversely, the focus is swept forward if the two images are above the horizontal distance bringing the object into focus. This system, of course, relies on there being enough ambient light for the system to work. Compact cameras and mobile sensors do not have the luxury of a discrete AF sensor, and use a contrast detection technique. The current mobile image sensor AF sweeps the lens through all positions and monitors the contrast of the imaged scene on the sensor in ISP – a computationally heavy and slow procedure. The algorithm in a CIS chooses the lens position with the greatest image contrast. Google's Project Tango extends the swept focus concept by implementing a CIS with

continuously sweeping focus in a mobile device, creating depth maps at 5 FPS with the processing cost of two GPU's. To reduce power and increase focus speeds in conventional imaging, AF pixels based on the phase detection principle have recently appeared in Sony's mirrorless DSLR and DSCs in the main sensor array and in both Sony and Samsung's mobile sensors [202]. Over the course of this research, ST's single point SPAD-based TOF range sensors (starting with the VL6180) have been adopted by LG in their G3 smartphone to provide fast AF assistance in competition to AF pixels.

A.5.4.7. 3D Vision Technology Summary

These different technologies are compared in terms of the applicability for a suite of 3D vision and depth sensing scenarios in Table A.5.1 in a simple 'yes, no or maybe' fashion. Green 'Y' denotes suitability for an application whereas a red 'N' denotes the opposite. A yellow 'maybe' highlights that there would be significant implementation challenges but would be possible. Table A.5.2 extends this by presenting a comparison reference table of system performance parameters against the competing technologies coloured in green, yellow and red to denote best to worst performance in that category.

		Low Res CIS		Stereo CIS		Structured Light	Light Field	Depth of Focus	Time of Flight	
Pixel		CIS							PMD	SPAD
Active Illumination		N	Y	N	Y	Y	N	N	Y	Y
Close Range (<10cm)	1D: Proximity	Y	Y	Maybe	Maybe	Y	N	N	Y	Y
	2D: Swipe Movement	Y	Y	Maybe	Maybe	Maybe	N	N	Y	Y
Medium Range (<1m)	1D: Auto-Focus Assist	N	N	N	N	N	N	Y	Y	Y
	2/3D: Hand & Finger Tracking	Maybe	Maybe	Y	Y	Y	Maybe	N	Y	Y
Medium to Far Range (>1m to 10m)	1D: Range Finder	N	N	N	N	N	N	N	Y	Y
	3D: Whole Body Detection / Tracking	N	N	Y	Y	Y	Maybe	N	Y	Y
Depth Map Capable		N	N	Y	Y	Y	Y	Y	Y	Y
Absolute Distance / Time Resolution		N	N	N	N	N	N	N	Y	Y
Single Photon Sensitivity		N	N	N	N	N	N	N	N	Y

Table A.5.1. A simple comparison table of the applicability between different CMOS sensor technologies for a range of 3D vision functions.

		Low Res CIS		Stereo CIS		Structured Light	Light Field	Depth of Focus	Time of Flight	
Pixel		CIS							PMD	SPAD
Active Illumination		N	Y	N	Y	Y	N	N	Y	Y
Area of Processing (ISP)		Med		High		Med	High	High	On-chip Pipeline	On-chip Pipeline
Power	Sensor	Low		Medium					High	Sys' Design & Light Dependent
	ISP	V. Low	Low	High (ISP)	High (ISP)	High (ISP)	High (ISP)	High (ISP)	Low	Low
	Illuminator Power	None	Low	None	Proportional to Distance Squared	High	None	None	Proportional to Distance Squared	
System Latency		Low		Med		Med	High	High	Low	Low
Ambient Light Performance	in Low Light	Fails	Good	Fails	Good	Good	Fails	Fails	Good	Good
	in Bright Indoors	Good	Good	Good	Good	Problematic	Good	Good	Affects Depth Accuracy	
	in Bright Outdoors	Good	Good	Good	Good	Fails	Good	Good	Worsening Depth Accuracy to Failure	Problematic w/ SPAD Paralysis
Spatial Resolution		Low Res (<8x8)		Scene Dependent		Pattern Dependent	MPixel	Low Res	QVGA to VGA	QVGA* to VGA (Roadmap)
Temporal Resolution		100's µs		ms (1 / integration time)					ns	ps
Depth Accuracy		None	None	µm to cm	µm to cm	mm to cm	cm	cm	mm to cm	mm to cm
Distance Range		<10cm	<10cm	10's cm to 10's m					1cm to 10's m	

Table A.5.2. Comparison table of 3D vision sensor technologies for a range of system performance parameters.

Appendix 6. Time-Domain Imaging Background

A.6.1. Time-Resolved Imaging Parameters

The following parameters are key performance indicators in TOF and time-domain imaging (2D/3D) and ranging (1D) systems. A brief definition of each follows.

A.6.1.1. Distance Range or Temporal Dynamic Range

The minimum to the maximum distance range that the TOF system can measure. In a time-domain measurement system, such as a TCSPC system, this is minimum to maximum time range the system can measure.

A.6.1.2. Distance Precision or Temporal Jitter

Standard deviation of repeated measurements. Does not include distance offset from measured distance value. In time-domain this is the system jitter. In distance, this is the distance precision. See temporal resolution for further discussion.

A.6.1.3. Distance Accuracy or Temporal Accuracy

The equivalent of data converter gain error in time or distance. The accuracy of the system is the deviation from real target distance or time to the mean distance or time obtained by the sensor over repeated measurements.

A.6.1.4. Spatial Resolution

The minimum angular or spatial separation that two neighbouring points at the same distance, that can be resolved by the imager.

A.6.1.5. Temporal Resolution

The minimum separation in time that two events can be resolved. For a single temporal event the temporal resolution is often confused with the time-domain jitter.

A.6.1.6. Sensor Frame Rate

The number of frames per second that the sensor can output. Not to be confused with the following parameter, TOF sensor acquisition rate. It is measured in frames per second (FPS).

A.6.1.7. Acquisition Rate

The output rate of 3D images; 3D indirect TOF cameras require four images (or time bins) to determine TOF. Some sensors capture the 4 images simultaneously and so the acquisition rate equals the frame rate. Others capture 2 images and so the 3D acquisition rate is half the frame rate, and some capture a single image quartering the acquisition rate versus sensor frame rate. This parameter in full systems also takes into account the system processing time.

A.6.1.8. Non-ambiguous Distance Range

TOF systems use an illumination signal with a certain repetition frequency. The maximum range that the system can measure without ambiguity is a full period of the illumination signal and is defined by

$$D_{max} = \frac{c}{2f_{illuminator}} \quad \text{(Eq.A.6.1)}$$

This distance is also referred to as the wraparound distance. Ambiguity in distance measurement can be seen in the following equation where a far target after the wraparound distance would report a close distance.

$$D_{reported}(\phi) = D(\phi + 2 . n . \pi) \quad \text{(Eq.A.6.2)}$$

where $\phi = \frac{1}{f_{illuminator}}$

A.6.1.9. Demodulation Contrast or Background Rejection

TOF systems capture and record reflected signals synchronised with a transmitted illumination signal in order to extract a phase shift (frequency domain) or time delay (time domain) to an object or target. The amplitude of the received signal is recorded into two, four or more time channels or bins. In order to resolve a target in the recorded signal, there must be a clear differential. Uncorrelated noise sources (such as ambient light) create a 'common mode' or common offset in all the time channels degrading the received differential signal. A key performance indicator in TOF systems is the ability to be unaffected by uncorrelated noise sources i.e. the

immunity of the system to background illumination. This can be quantified by demodulation contrast as shown in the following expression:

$$\chi = \frac{Amplitude}{Offset} = \frac{Differential}{Common\ Mode} \quad \text{(Eq.A.6.3)}$$

where in a sinusoidal system the offset is only from the recorded signal and the expression is unity. In a pulsed system the expression is similar where amplitude and offset are normalised:

$$\chi = \frac{Amplitude}{1 + Offset} = \frac{Differential}{1 + Common\ Mode} \quad \text{(Eq.A.6.4)}$$

Background rejection in many published works is defined in 'lux' unit of luminance. Although, it is important to note that the 'lux' unit is weighted to the perception of the human eye and not appropriate for TOF sensors working in NIR.

A.6.1.10. Dynamic Range

The dynamic range is the minimum to the maximum light level that the sensor can record. It is a key parameter as the intensity of the illumination reflection varies over a large range as it is proportional to the distance and lens F number squared:

$$Reflection\ Intensity \propto \left(\frac{1}{D\ .\ F\#}\right)^2 \quad \text{(Eq.A.6.5)}$$

For very short distances the reflection is intense, and at long distances the reflection captured by the sensor can often fall below one photon on average per laser repetition [203]. Therefore high QE and FF and a low F number lens are central to achieving the greatest possible dynamic range.

A.6.2. TOF Illumination Techniques

There are three primary TOF illumination modulation techniques used in TOF imaging and ranging: amplitude, frequency and temporal modulation. Amplitude modulation has a single frequency and varies the intensity of the illumination. Frequency modulation varies both frequency and intensity. Temporal modulation varies signals in the time-domain. There are hybrids of these also, for example amplitude modulation (AM) may be operated with selective frequency switching and frequency modulation (FM) may be operated with selective amplitude switching.

Each of the three main categories can either be continuous or discontinuous. The former, 'continuous', refers to an illumination signal which has 50:50 duty cycle. The latter, describes signals which pulse or transmit for an 'on' period then cease for an 'off' period having a duty cycle less than 50:50. DTOF uses narrow pulse illumination that falls under the category of discontinuous temporal modulation. A non-exhaustive list, of the ITOF illumination techniques, is as follows:

- Amplitude modulation (AM)
    - AM continuous wave (AMCW)
    - AMCW with selective frequency switching.
    - AM discontinuous wave (AMDW) – not yet described in the literature.
- Frequency modulation (FM)
    - FM continuous wave.
    - FMCW with selective amplitude switching.
    - FM discontinuous wave (FMDW) – not yet described in the literature.
- Time gating or temporal modulation (TM)
    - Square wave continuous wave is both AMCW and temporal modulation CW (TMCW).
    - TMCW with selective amplitude and, or frequency switching.
    - TM discontinuous pulse (TMDP) – DTOF pulsed laser illumination falls under this category.

There are two illumination techniques for extending the non-ambiguous distance range of a TOF sensor:

- Non-ambiguous distance range extension
- Discontinuous illumination.

A.6.2.1. Non-Ambiguous Distance Range Extension

Buttgen and Seitz [204] detail an AMCW approach using two frequencies f_1 and f_2, where the maximum range is given by the difference between the two:

$$D = \frac{1}{4\pi \,.\, c \,.\, (f_2 - f_1)}$$

(Eq.A.6.6)

A.6.2.2. Range Extension through Discontinuous Illumination

The second technique is used in pulsed illumination for DTOF, using a low duty-cycle illumination pulse. Here this technique is described generally as discontinuous wave (DCW) illumination. Table A.6.3 describes the operational differences between continuous wave (CW) and DCW illumination in terms of a number of parameters. In discontinuous illumination an off or disabled period is used after one repetition of the illuminator signal hence lowering the duty cycle of discontinuous illumination below the CW 50:50. This is achieved, in order to mitigate TOF wrapping and to extend the non-ambiguous distance range as illustrated in Figure A.6.1 (whereby a far target at 'the wrapping distance' appears as a faint close target). The real frequency of the DCW is the reciprocal of the period of the illumination wave (T_{PW}) plus the off period (T_{OFF}). On the other hand, the effective illumination frequency of the system is solely the inverse of the illumination period. The power of the illumination is less in a DCW system although it may take longer for a discontinuous system to attain the same signal within an integration period.

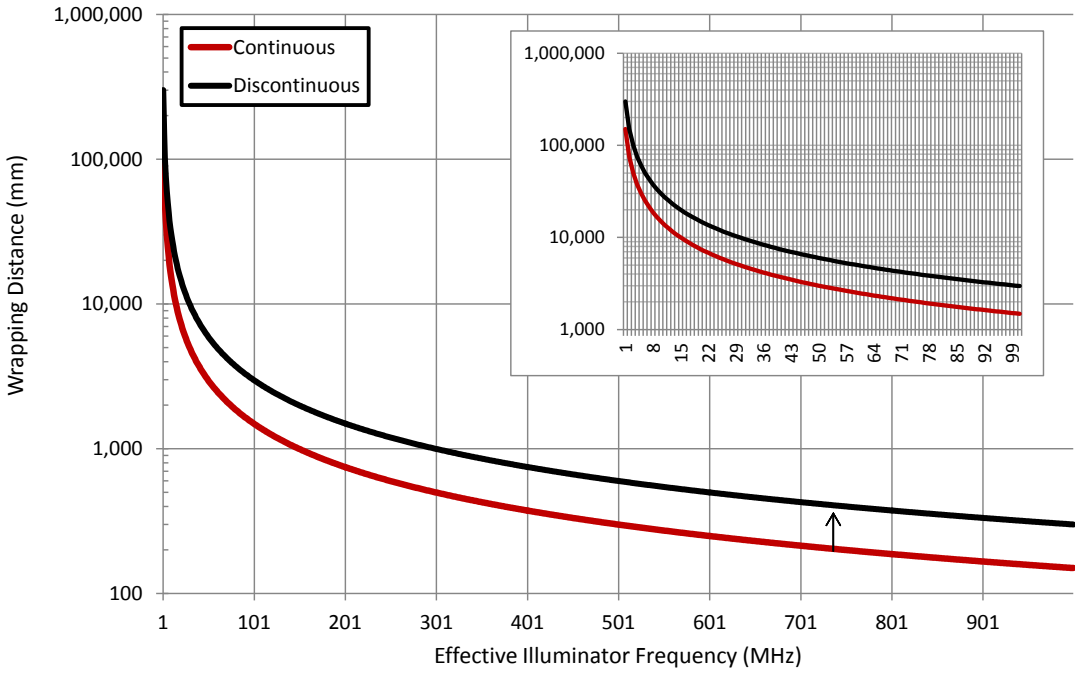


Figure A.6.1. Graph of non-ambiguous distance versus effective illuminator frequency for CW versus DCW illumination with inset zoom-in of 1 to 100MHz frequency.

A.6.2.3. Illuminator Pulse Width and Duty Cycle.

In contrast to sinusoidal illumination, a pulsed ITOF system can be configured to have <50% duty cycle illumination which has two distinct advantages. Firstly, in two or four bin ITOF shortening the illumination pulse provides greater distance precision at the cost of shorter dynamic range. Secondly reducing the illuminator repetition frequency provides a reduction in optical power and increases the TOF wrapping distance.

Table A.6.1 illustrates a scenario varying illuminator pulse width against a fixed illuminator 50MHz repetition rate. A shorter pulse width is the equivalent of a greater sinusoidal illuminator frequency. Much greater than 100MHz is difficult to realise in a high power LED (as found in many commercial TOF systems) and the first five table entries are underlined where a short pulse laser illuminator combined with a SPAD-based TOF system has an advantage over competitor LED-based systems.

Illuminator Pulse Width (ns)	Equivalent Sinusoidal Illuminator Frequency (MHz)	TOF Dynamic Range (m)	Average Optical Power as a % of Sinusoidal Equivalent.
1	500.0	0.15	10%
2	250.0	0.30	20%
3	166.7	0.45	30%
4	125.0	0.60	40%
5	100.0	0.75	50%
6	83.3	0.90	60%
7	71.4	1.05	70%
8	62.5	1.20	80%
9	55.6	1.35	90%
10	50.0	1.50	100%

Table A.6.1. Illuminator pulse width varied against a fixed illuminator 50MHz repetition rate. Underlined entries indicate where a SPAD-based pulsed ITOF system has an advantage over sinusoidal ITOF systems.

Table A.6.2 illustrates a second scenario varying illuminator repetition frequency against a fixed 5ns pulse width. By lowering the duty cycle of the illumination under a 1:1 (50%:50%) ratio, the wrapping distance (the equivalent dynamic range) increases, but importantly there is a region of distance that can be time-gated off after the 1.5m dynamic range of the 5ns pulse width. This region from 1.5m to the wrapping distance (represented in the equation is an effective 'dead zone' to the TOF camera and inherently removes out of range targets. Four entries are underlined indicating a compromise between repetition frequency and dead zone length.

Illuminator Repetition Frequency (MHz)	Equivalent Dyn. Range or Wrapping Distance (m)	Illuminator Duty Cycle	Dead Zone from 1.5m Dyn. Range to Wrapping Distance (m)	Average Optical Power as a % of 100MHz equivalent.
100	1.5	1:1 (50% : 50%)	0	100%
66.6	2.3	1:1.5	0.75	66%
50	3.0	1:2	1.50	50%
33.3	4.5	1:3	3.00	33%
25	6.0	1:4	4.50	25%
12.5	12	1:8	10.5	12.5%
6.25	24	1:16	22.5	6.3%
3.125	48	1:32	46.5	3.1%

Table A.6.2. Illuminator repetition rate varied against fixed 5ns illuminator pulse width. Underlined entries indicate an effective compromise between high enough repetition frequency and dead zone length.

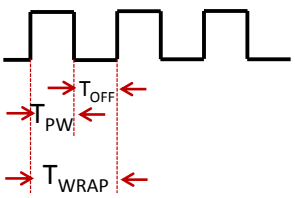
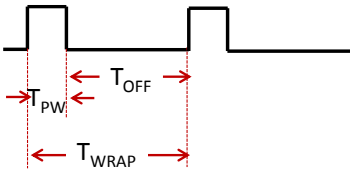
	Continuous Wave (CW) Illumination	Discontinuous Wave (DCW) Illumination
Waveform Diagram		
Duty Cycle	50:50 : $T_{PW} = T_{OFF}$	Less than 50:50 : $T_{PW} < T_{OFF}$
Real Illuminator Frequency	$f_{real} = f_{eff} = \frac{1}{2T_{PW}}$ No independent control.	$f_{real} = \frac{1}{T_{WRAP}} = \frac{1}{T_{PW} + T_{OFF}}$ Real frequency independent of effective frequency.
Effective Illuminator Frequency		$f_{eff} = \frac{1}{2T_{PW}}$ High effective frequency with short pulse width.
Wrapping Time	$T_{WRAP} = 2 \, . \, T_{PW}$	$T_{WRAP} = T_{PW} + T_{OFF}$
Non-Ambiguous Range Distance	$Z_{WRAP} = c \, . T_{WRAP} = c \, . \, 2T_{PW}$ Trade-off exists between precision (greater at high frequency) and wrapping distance (higher for lower frequencies)	$Z_{WRAP} = c \, . T_{WRAP}$ $Z_{WRAP} = c \, . \, (T_{PW} + T_{OFF})$ Longer wrap distance, keep precision and high effective frequency.
Illuminator Power	$P \propto f_{eff} \propto \frac{1}{2 \, T_{PW}}$ Dependent on frequency	$P \propto f_{eff} \propto \frac{1}{T_{PW} + T_{OFF}}$ Proportional to duty cycle and pulse width.

Table A.6.3. Continuous versus discontinuous ITOF illumination.

A.6.2.4. Choice of Illumination for SPAD-based ITOF Imaging

AMCW is well described in the literature and are the basis of all ITOF image sensors on the market. SPAD-based image sensors have been described using AMCW in [36], [53]. Yet, where SPADs have an advantage is using the detector's picosecond temporal resolution combined with integration on chip of short nano-second time gating permits higher effective illuminator frequency (>100 MHz) that is not achievable with photo-demodulation devices as described in the following section. To increase the signal to noise ratio of a TOF system, assuming we are limited by average optical power for laser safety regulations, powerful short pulses with low duty cycle have considerably higher signal for a short duration than a sinusoidal source with the same average power. For these reasons, temporal modulation with pulsed illumination is preferred. However, to increase SNR by reducing noise, the solution lies in the pixel.

A.6.3. ITOF Pixels

This section provides a brief overview of the literature on photo diode and photo-gate time-domain image sensor pixels.

A.6.3.1. ITOF Charge Domain Pixels

To demodulate a frequency domain signal or to capture fast optical phenomena, shuttering can be applied in the charge domain to capture and shift charge only within a defined short window. This section details a number of charge domain photo-demodulation ITOF and temporal imaging pixel designs. The generalised operation of a photo-gate demodulator is illustrated in Figure A.6.2 (a) and a photo-mixing device in Figure A.6.2 (b).

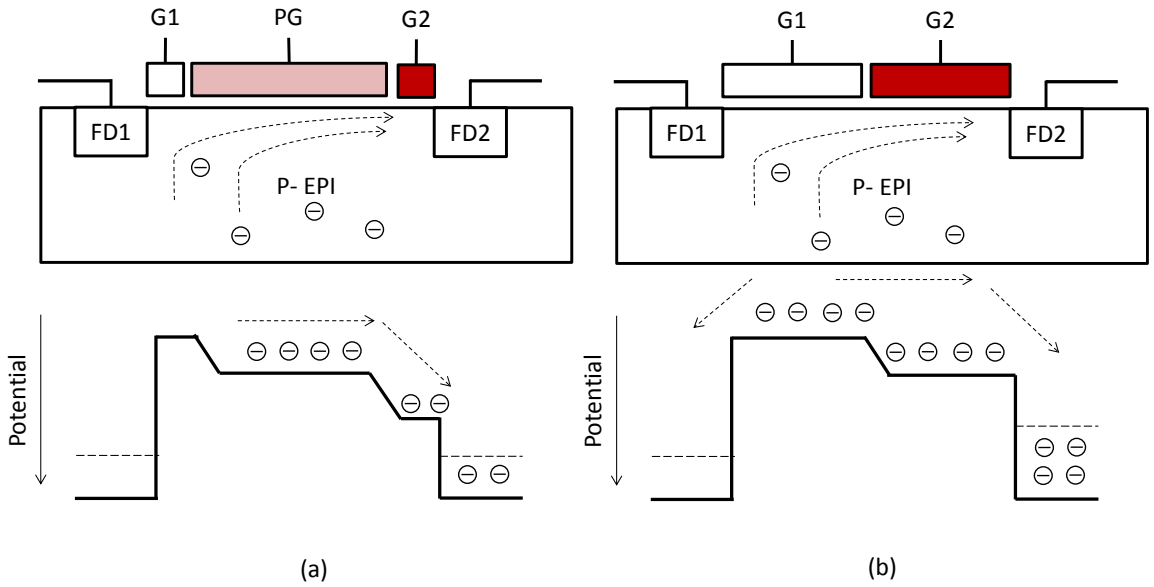


Figure A.6.2. Diagram reproduced from [65]**. (a) Photo-gate demodulator structure. (b) Photo-mixing device structure.**

Spirig, Seitz, et al. (then PSI now CSEM) were the first to present a CCD photogate-based demodulation pixel in 1995 referred to as a 'Lock-In Pixel' [205], [206]. They implemented a 3x3 array of a pixel with four transfer gates and a 'dump' gate (time-domain equivalent of a CIS anti-bloom device). Seitz then implemented the first imaging array of 64x25 single-tap photo-gate pixels in 2001 attaining a distance precision of 3.84cm (equivalent of 260ps) using a pixel of 65x21µm with 22% FF [207].

Between 2002 and 2006, three companies released their initial TOF cameras to the market. PMDTec's pixel 'photonic mixing device' (PMD) is an optimised two-tap photo-gate structure that was first presented by Schwartz (University of Seigen) in 1997 [208] and improved to 50µm PP and a demodulation contrast of 50% and the ability to resolve >100MHz illumination in 2005 [204], [209]. Mesa Imaging's pixel was published in 2004 at 50µm PP with 17% FF and 40% reported demodulation contrast by Buttgen et al. [210].

Softkinetic, the third of the startup companies, implements an alternative structure which creates a drift field using a current flow entitled a current assisted photo demodulator (CAPD). In the initial work [211], the current consumption per 30x25µm pixel is excessive at 1.3mA. This current consumption has been addressed in separate works by Stoppa et al. [212] and Dalla Betta et al. [213]. Yet, the current consumption remains, which eventually limits the scalability to megapixel arrays.

In 2010, the first 10µm PP CIS implementations were presented at ISSCC. Stoppa et al. (FBK) demonstrated a PIN photo-diode based demodulator pixel with two taps at 24% fill-factor operational up to a maximum of 50MHz [214]. Samsung presented two RGBZ imagers with 10µm and 14 µm pitch [173]. Kawahito et al. have

recently published a range of works on two-tap photo-diode based pixels for TOF and FLIM under the banner of 'lateral electric field modulation' (LEFM) [215], [216].

However, the process of moving electric charge in silicon is not instantaneous there is a fundamental limit of electron diffusion time. Drift fields, applied electric fields, large potential differences between charge storage nodes and high conversion factor (CVF) all contribute to decreasing this temporal delay in charge transfer yet the limit remains. This concept is illustrated in Figure A.6.3, that short nanosecond pulse cannot be fully demodulated due to e- drift time and so high frequency TOF illuminator frequencies cannot be fully demodulated (the demodulation contrast decreases for increasing illuminator frequency). Charge-domain pixels can only transfer charge to one in-pixel location (storage node) at a time and so are limited to serial non-overlapping exposure windows.

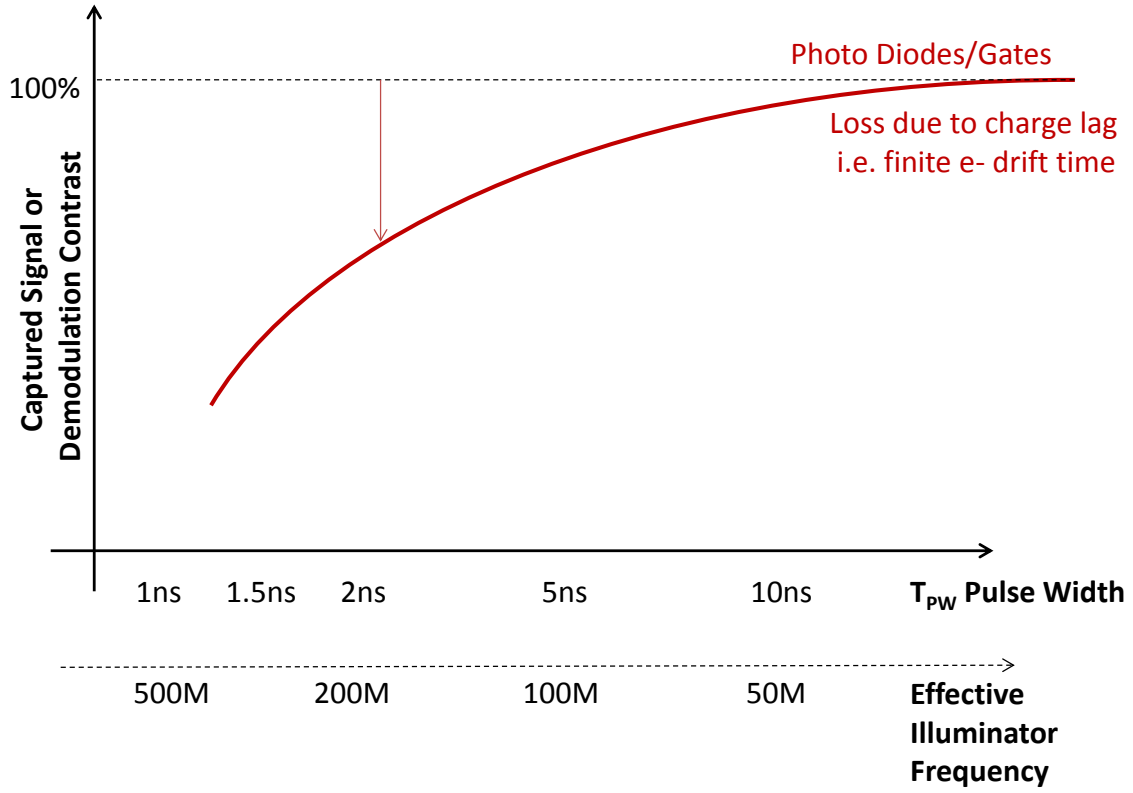


Figure A.6.3. Illustrative trend of illuminator pulse width versus demodulation contrast for photo-gate and photo-diode based time-domain pixels.

A notable concept to minimise time-gate row driver mismatch was presented in 2013 by Yasutomi, Kawahito et al. whereby DAC controlled 'trimmable' delay cells on each row driver were implemented [216]. Such digitally assisted analogue techniques will become critical in all SPAD sensors design for achieving picosecond mismatch between channels, rows or columns as the move is made to smaller process geometries.

A.6.3.2. ITOF Voltage Domain Pixels.

A number of photo-diode pixels have been described in the literature that use a fast electronic shutter with voltage domain in-pixel integration to capture ITOF. The advantage of these pixels is they can perform enhanced timing functions over charge domain pixels, and can be implemented in standard and imaging CMOS. The disadvantage is the addition of CMOS electronics decreases fill factor and adds noise: both factors decreasing SNR.

Jeremias et al. (Fraunhofer) presented the first voltage domain ITOF pixel in a 32x2 linear array in 2001 at ISSCC in [217] (the first time-domain imaging array shown at ISSCC). Elkhalili et al. expanded the original sensor in 2006 to 64x8 array with balanced clock routing, nanosecond time gating (30 ns min) but large pixel at 300x130µm [218] . They report '100% fill factor' of only the 130x130µm photo-diode, yet when the pixel electronics is taken into account it is estimated at 43%. The pixel circuit, which performs background subtraction through switched capacitor operation, is shown in Figure A.6.4. The front end 1x gain buffer does not provide any photo-diode gain and by examination attenuates the input signal through the parallel capacitance of the photodiode and 'Cs'. The power consumption is high at 3W for the latter 2006 sensor.

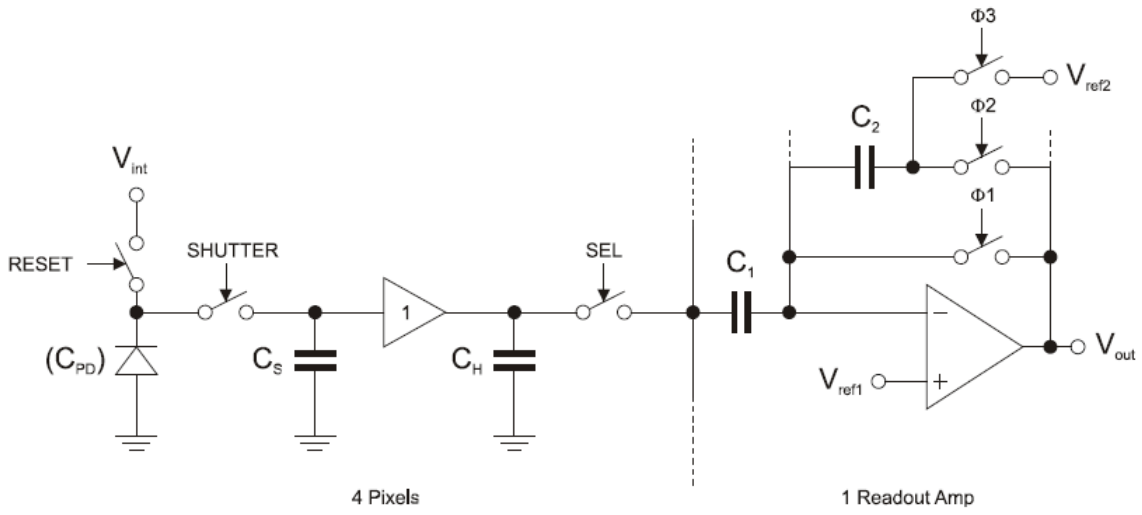


Figure A.6.4. Jeremias and Elkhalili (Fraunhofer) voltage domain ITOF pixel. Reproduced with permission from [219]**.**

Stoppa et al. (FBK) demonstrated a pseudo-differential pixel architecture that performed background subtraction through switched capacitor action. A 16x16 array was implemented in ESSCIRC 2004 and a 32x32 array at IISW 2005. The 81x81µm pixel circuit is shown in Figure A.6.5, attains a 20% fill factor. The front end amplifier provides 8x gain at the front end.

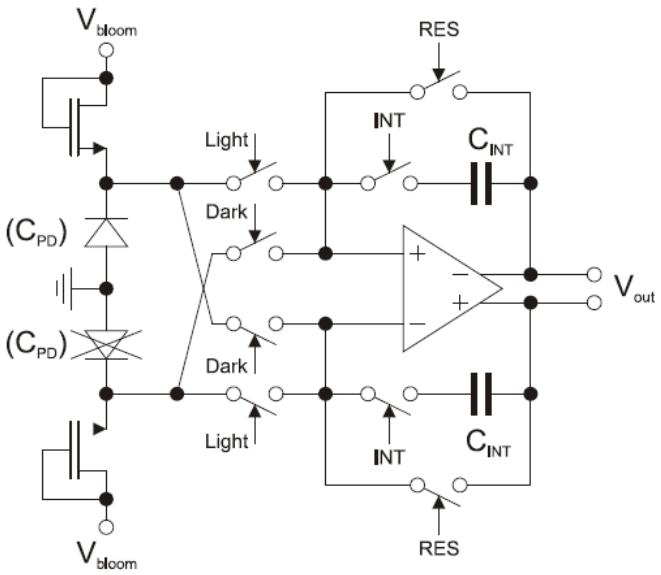


Figure A.6.5. Stoppa (FBK) pseudo-differential voltage domain ITOF pixel. Reproduced with permission from [219]**.**

Zach et al. (University of Vienna) implemented a switched capacitor differential pair pixel design performing background subtraction in a 125x125µm pixel with the highest 66% fill factor of a voltage domain pixel. It was implemented in a 32x2 line sensor in 2009 [220], [221] and a 16x16 array in 2010 in [222], [223] reporting suppression of 150 to 180klux of ambient light. There is no amplification of the input signal in these pixel designs. On the other hand this pixel must continuously operate and the two windows must be non-overlapping.

Perenzoni et al. (FBK) implemented a 160x120 array in 2011 [224], [225]. At 29.1µm pitch at 34% fill factor it lowered the pixel area of the previous 85µm by 88% and increased the fill factor. The front end amplifier has 8x gain and the pixel was trialled up to 128x integration with background subtraction up to 20klux. The sensor had 314mW power consumption. The switched capacitor pixel is shown in Figure A.6.6.

These voltage domain pixels can provide background subtraction functionality and furthermore could implement different timing schemes such as overlapping time-gate windows that charge domain pixels cannot implement. However, the inclusion of many transistors lowers the fill factor, limits the pixel pitch and introduces noise sources such as cyclic charge injection and kT/C noise on integration capacitors. Chip stacking technology would facilitate placing the pixel electronics on a lower IC under the photo-diode to mitigate the low fill factor.

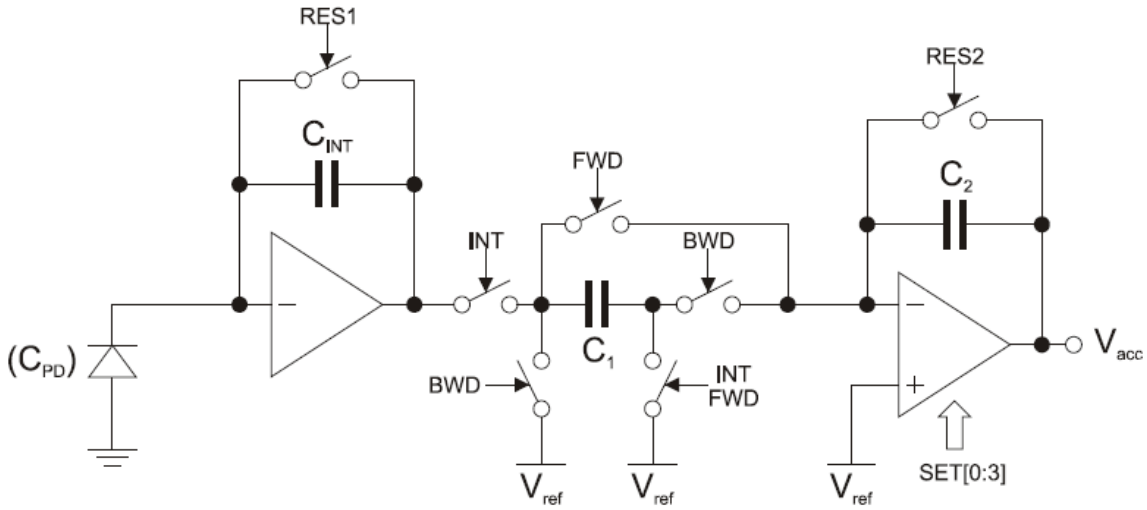


Figure A.6.6. Perenzoni voltage domain ITOF pixel circuit. Reproduced with permission from [219]**.**

Appendix 7. Noise Sources Reference

Figure A.7.1 illustrates, for reference, the principal noise sources from the SPAD, the pixel counter and source follower and through the analogue readout chain.

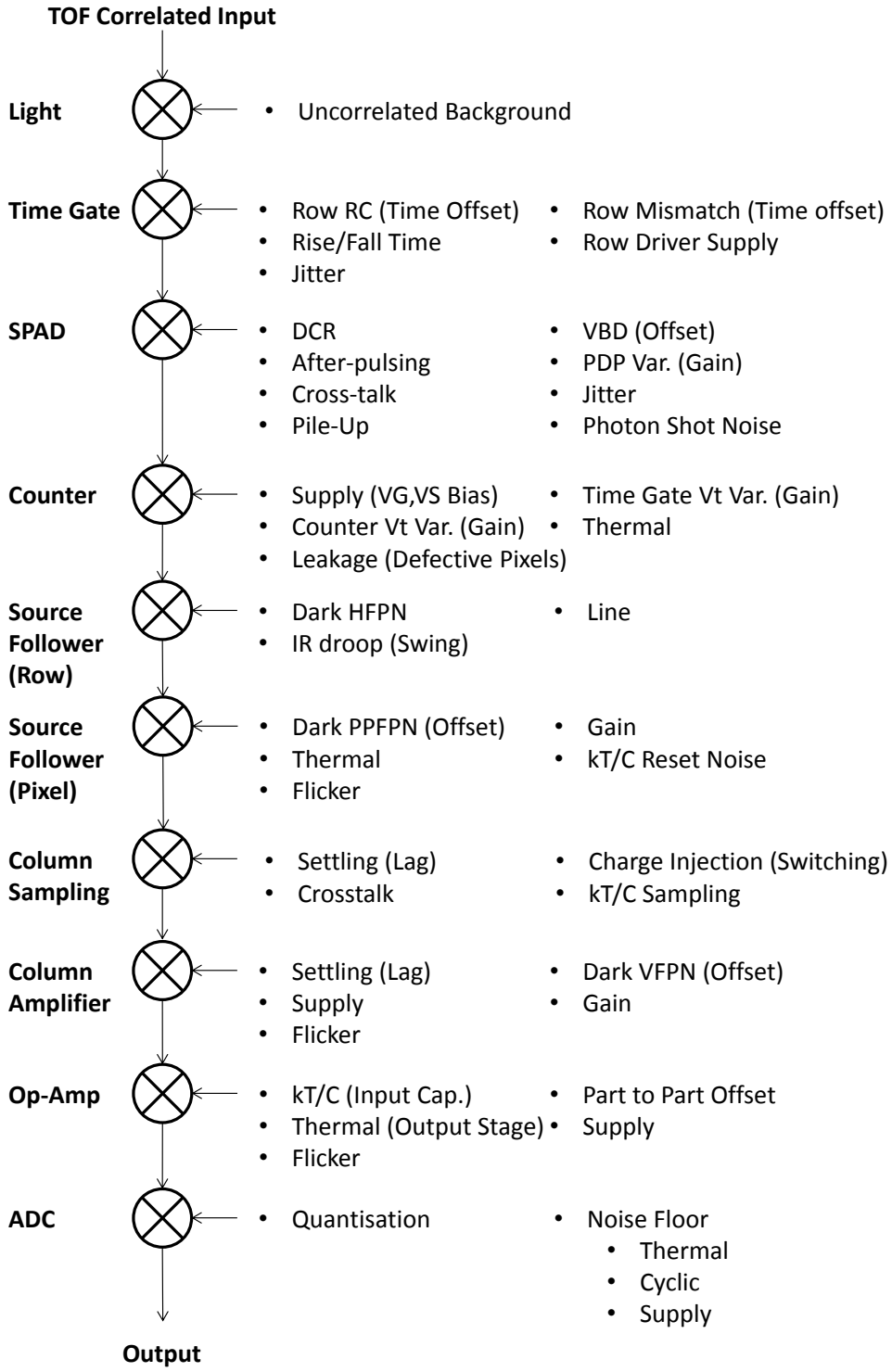


Figure A.7.1. SPAD image sensor noise sources in the style of Theuwissen's CIS noise source diagram [226].

References

[1] N. A. W. Dutton, L. A. Grant, and R. K. Henderson, "9.8μm SPAD-based Analogue Single Photon Counting Pixel with Bias Controlled Sensitivity," in *International Image Sensors Workshop*, 2013.

[2] N. A. W. Dutton, L. Parmesan, A. J. Holmes, L. A. Grant, and R. K. Henderson, "320x240 Oversampled Digital Single Photon Counting Image Sensor," in *VLSI Symposium*, 2014.

[3] N. Dutton, J. Vergote, S. Gnecchi, L. Grant, D. Lee, S. Pellegrini, B. Rae, and R. Henderson, "Multiple-event direct to histogram TDC in 65nm FPGA technology," in *2014 10th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME)*, 2014, pp. 1–5.

[4] L. Parmesan, N. A. W. Dutton, N. J. Calder, A. J. Holmes, L. A. Grant, and R. K. Henderson, "A 9.8 μm Sample and Hold Time to Amplitude Converter CMOS SPAD Pixel," in *ESSDERC 2014*, 2014.

[5] N. A. W. Dutton, S. Gnecchi, L. Parmesan, A. J. Holmes, B. Rae, L. A. Grant, and R. K. Henderson, "A Time Correlated Single Photon Counting Sensor with 14GS/s Histogramming Time to Digital Converter," in *IEEE International Solid-State Circuits Conference - ISSCC Digest of Technical Papers*, 2015.

[6] N. A. W. Dutton, L. Parmesan, S. Gnecchi, I. Gyongy, N. J. Calder, B. R. Rae, L. A. Grant, and R. K. Henderson, "Oversampled ITOF Imaging Techniques using SPAD-based Quanta Image Sensors," in *International Image Sensor Workshop*, 2015.

[7] L. Parmesan, N. A. W. Dutton, N. J. Calder, L. A. Grant, and R. K. Henderson, "A 256x256 SPAD array with in-pixel Time to Amplitude Conversion for Fluorescence Lifetime Imaging Microscopy," in *International Image Sensor Workshop*, 2015.

[8] I. Gyongy, N. A. W. Dutton, L. Parmesan, A. Davies, R. Saleeb, R. Duncan, C. Rickman, P. Dalgarno, and R. K. Henderson, "Bit-plane Processing Techniques for Low-Light, High Speed Imaging with a SPAD-based QIS," in *International Image Sensor Workshop*, 2015.

[9] O. Almer, N. A. W. Dutton, T. Al Abbas, S. Gnecchi, and R. K. Henderson, "4-PAM visible light communications with a XOR-tree digital silicon photomultiplier," in *2015 IEEE Summer Topicals Meeting Series (SUM)*, 2015, pp. 41–42.

[10] S. Gnecchi, N. A. W. Dutton, L. Parmesan, B. R. Rae, S. Pellegrini, L. A. Grant, and R. K. Henderson, "Gphotons/s TCSPC with XOR Digital Silicon Photomultipliers," in *Single Photon Workshop*, 2015.

[11] S. Gnecchi, N. A. W. Dutton, L. Parmesan, B. R. Rae, S. Pellegrini, S. J. McLeod, L. A. Grant, and R. K. Henderson, "A Comparative Analysis between OR-based and XOR-based Digital Silicon Photomultipliers for PET," in *IEEE Nuclear Science Symposium*, 2015.

[12] O. Almer, D. Tsonev, N. A. W. Dutton, T. Al Abbas, S. Videv, S. Gnecchi, H. Haas, and R. K. Henderson, "A SPAD-based Visible Light Communications Receiver Employing Higher Order Modulation," in *IEEE Globecom*, 2015.

[13] S. P. Poland, N. Krstajić, J. Monypenny, S. Coelho, D. Tyndall, R. J. Walker, V. Devauges, J. Richardson, N. A. W. Dutton, D. D.-U. Li, K. Suhlinge, M. Parsons, T. Ng, R. K. Henderson, and S. M. Ameer-Beg, "A Time-Tesolved Multifocal Multiphoton Microscope for High Speed FRET imaging In Vivo," in *Focus on Microscopy*, 2014.

[14] R. Henderson, N. Krstajić, N. Dutton, and R. Walker, "CMOS Single Photon Image Sensors For Microscopy," in *Microscience and Microscopy Congress (MMC) 2014*, 2014.

[15] S. P. Poland, N. Krstajić, J. Monypenny, S. Coelho, D. Tyndall, R. Walker, V. Devauges, J. A. Levitt, N. Dutton, T. Ng, R. Henderson, and S. Ameer-Beg, "A time-resolved multifocal multiphoton microscope for high speed fret imaging in vivo," in *Microscience and Microscopy Congress (MMC) 2014*, 2014.

[16] S. P. Poland, N. Krstajić, J. Monypenny, S. Coelho, D. Tyndall, R. J. Walker, V. Devauges, J. Richardson, N. Dutton, P. Barber, D. D.-U. Li, K. Suhling, T. Ng, R. K. Henderson, and S. M. Ameer-Beg, "A high speed multifocal multiphoton fluorescence lifetime imaging microscope for live-cell FRET imaging," *Biomed. Opt. Express*, vol. 6, no. 2, p. 277, Jan. 2015.

[17] N. A. W. Dutton, I. Gyongy, L. Parmesan, S. Gnecchi, N. Calder, B. R. Rae, S. Pellegrini, L. A. Grant, and R. K. Henderson, "A SPAD-Based QVGA Image Sensor for Single-Photon Counting and Quanta Imaging," *IEEE Trans. Electron Devices*, vol. Pre-Public, no. 99, pp. 1–8, 2015.

[18] S. Gnecchi, N. A. W. Dutton, L. Parmesan, B. R. Rae, S. Pellegrini, S. J. McLeod, L. A. Grant, and R. K. Henderson, "Digital Silicon Photomultipliers With OR/XOR Pulse Combining Techniques," *IEEE Trans. Electron Devices*, vol. 63, no. 3, pp. 1105–1110, Mar. 2016.

[19] S. Gnecchi, N. Dutton, L. Parmesan, L. Grant, and R. Henderson, "Analysis of Photon Detection Efficiency and Dynamic Range in SPAD based Visible Light Receivers," *J. Light. Technol.*, vol. PP, no. 99, pp. 1–1, 2016.
[20] N. A. W. Dutton, I. Gyongy, L. Parmesan, and R. K. Henderson, "Single Photon Counting Performance and Noise Analysis of CMOS SPAD-based Image Sensors," *Sensors (Basel).*, no. Special Issue, pp. 1–17, 2016.
[21] N. A. W. Dutton, "Real-Time Processing Method and System," US 20120162632 A128-Jun-2012.
[22] N. A. W. Dutton, "Error correction in thermometer codes," US 8432304 B230-Apr-2013.
[23] N. A. W. Dutton, "Circuit for combining signals," US 9007118 B214-Apr-2015.
[24] N. A. W. Dutton and R. K. Henderson, "Improvements in time of flight pixel circuits," EP 2728373 A1, US2014012465207-May-2014.
[25] N. A. W. Dutton and R. K. Henderson, "Pixel circuit with capacitor discharge indicative of number of events," US 20140124653 A108-May-2014.
[26] N. A. W. Dutton, S. Gnecchi, and R. K. Henderson, "Time to digital converter and applications thereof," US 20150041625 A112-Feb-2015.
[27] G. P. Weckler, "A silicon photodevice to operate in a photon flux integrated mode," in *1965 International Electron Devices Meeting*, 1965, vol. 11, pp. 38–39.
[28] P. J. W. Noble, "Self-scanned silicon image detector arrays," *IEEE Trans. Electron Devices*, vol. 15, no. 4, pp. 202–209, Apr. 1968.
[29] M. F. . Tompsett, "Lecture to International Image Sensors Workshop." Salt Lake City, Utah, 2013.
[30] W. Becker, *Advanced Time-Correlated Single Photon Counting Techniques*. Springer, 2005.
[31] T. Spirig, P. Seitz, O. Vietze, and F. Heitger, "The lock-in CCD-two-dimensional synchronous detection of light," *IEEE J. Quantum Electron.*, vol. 31, no. 9, pp. 1705–1708, 1995.
[32] B. F. Aull, A. H. Loomis, D. J. Young, R. M. Heinrichs, B. J. Felton, P. J. Daniels, and Deborah J. Landers, "Geiger-Mode Avalanche Photodiodes for Three Dimensional Imaging," *Lincoln Lab. J.*, vol. 13, no. 2, pp. 335–350, 2002.
[33] A. Rochas, A. R. Pauchard, P.-A. Besse, D. Pantic, Z. Prijic, and R. S. Popovic, "Low-noise silicon avalanche photodiodes fabricated in conventional CMOS technologies," *IEEE Trans. Electron Devices*, vol. 49, no. 3, pp. 387–394, Mar. 2002.
[34] C. L. Niclass, A. Rochas, P. A. Besse, and E. Charbon, "A CMOS single photon avalanche diode array for 3D imaging," *IEEE Int. Solid-State Circuits Conf. - ISSCC Dig. Tech. Pap.*, 2004.
[35] C. Niclass, M. Soga, H. Matsubara, S. Kato, and M. Kagami, "A 100-m Range 10-Frame/s 340x96-Pixel Time-of-Flight Depth Sensor in 0.18-µm CMOS," *IEEE J. Solid-State Circuits*, vol. 48, no. 2, pp. 559–572, Feb. 2013.
[36] C. Niclass and E. Charbon, "A single photon detector array with 64-64 resolution and millimetric depth accuracy for 3D imaging," in *IEEE International Solid-State Circuits Conference - ISSCC Digest of Technical Papers*, 2005, pp. 364–366.
[37] D. Stoppa, F. Borghetti, J. Richardson, R. Walker, L. Grant, R. K. Henderson, M. Gersbach, and E. Charbon, "A 32x32-pixel array with in-pixel photon counting and arrival time measurement in the analog domain," in *Proceedings of the European Solid-State Circuits Conference (ESSCIRC)*, 2009, pp. 204–207.
[38] M. Gersbach, Y. Maruyama, E. Labonne, J. Richardson, R. Walker, L. Grant, R. Henderson, F. Borghetti, D. Stoppa, and E. Charbon, "A Parallel 32x32 Time-To-Digital Converter Array Fabricated in a 130 nm Imaging CMOS Technology," *Proc. Eur. Solid-State Circuits Conf.*, pp. 196–199, 2009.
[39] R. J. Walker, "A Fully Digital, Phase-Domain $\Delta\Sigma$ 3D Range Image Sensor in 130nm CMOS Imaging Technology," 2011.
[40] J. Richardson, R. Walker, L. Grant, D. Stoppa, F. Borghetti, E. Charbon, M. Gersbach, and R. K. Henderson, "A 32x32 50ps Resolution 10 bit Time to Digital Converter Array in 130nm CMOS for Time Correlated Imaging," in *IEEE Custom Integrated Circuits Conference*, 2009, pp. 77–80.
[41] R. J. Walker, L. H. C. Braga, A. T. Erdogan, L. Gasparini, L. A. Grant, R. K. Henderson, N. Massari, M. Perenzoni, and D. Stoppa, "A 92k SPAD Time-Resolved Sensor in 0.13µm CIS Technology for PET/MRI Applications," in *International Image Sensor Workshop*, 2013.
[42] G. Gariepy, N. Krstajić, R. Henderson, C. Li, R. R. Thomson, G. S. Buller, B. Heshmat, R. Raskar, J. Leach, and D. Faccio, "Single-photon sensitive light-in-fight imaging," *Nat. Commun.*, vol. 6, p. 6021, Jan. 2015.

[43] P. Seitz and A. Theuwissen, *Single Photon Imaging*, 1st ed. Springer, 2011.

[44] E. R. Fossum, "Quanta Image Sensor : Possible paradigm shift for the future," in *Proceedings of Image Sensors Europe 2012*, 2012, pp. 1–37.

[45] E. R. Fossum, "Gigapixel Digital Film Sensor (DFS) Proposal," in *Nanospace Manipulation of Photons and Electrons for Nanovision Systems, Proceedings of.*, 2005.

[46] J. Nakamura, Ed., *Image Sensors and Signal Processing for Digital Still Cameras*, First. CRC Press, 2006.

[47] R. K. Henderson, E. A. G. Webster, R. Walker, J. A. Richardson, and L. A. Grant, "2010 International Electron Devices Meeting," in *2010 IEEE International Electron Devices Meeting (IEDM)*, pp. 14.2.1–14.2.4.

[48] E. A. G. Webster, R. J. Walker, R. K. Henderson, and L. A. Grant, "A silicon photomultiplier with >30% detection efficiency from 450–750nm and 11.6μm pitch NMOS-only pixel with 21.6% fill factor in 130nm CMOS," in *2012 Proceedings of the European Solid-State Device Research Conference (ESSDERC)*, 2012, pp. 238–241.

[49] "Canesta's Latest 3D Sensor - 'Cobra,'" 2010. [Online]. Available: https://www.youtube.com/watch?v=5_PVx1NbUZQ. [Accessed: 08-Apr-2014].

[50] W. Kim, W. Yibing, I. Ovsiannikov, S. Lee, Y. Park, C. Chung, and E. Fossum, "A 1.5Mpixel RGBZ CMOS image sensor for simultaneous color and range image capture," in *2012 IEEE International Solid-State Circuits Conference*, 2012, pp. 392–394.

[51] S.-J. Kim, B. Kang, J. D. K. Kim, K. Lee, C.-Y. Kim, and K. Kim, "A 1920×1080 3.65μm-pixel 2D/3D image sensor with split and binning pixel structure in 0.11pm standard CMOS," in *IEEE International Solid-State Circuits Conference - ISSCC Digest of Technical Papers*, 2012, pp. 396–398.

[52] D. Stoppa, N. Massari, L. Pancheri, M. Malfatti, M. Perenzoni, and L. Gonzo, "A Range Image Sensor Based on 10um Lock-In Pixels in 0.18um CMOS Imaging Technology," *IEEE J. Solid-State Circuits*, vol. 46, no. 1, pp. 248–258, Jan. 2011.

[53] R. J. Walker, J. A. Richardson, and R. K. Henderson, "A 128×96 pixel event-driven phase-domain ΔΣ-based fully digital 3D camera in 0.13μm CMOS imaging technology," in *2011 IEEE International Solid-State Circuits Conference*, 2011, pp. 410–412.

[54] C. Veerappan, J. Richardson, R. Walker, D.-U. Li, M. W. Fishburn, Y. Maruyama, D. Stoppa, F. Borghetti, M. Gersbach, R. K. Henderson, and E. Charbon, "A 160×128 single-photon image sensor with on-pixel 55ps 10b time-to-digital converter," in *2011 IEEE International Solid-State Circuits Conference*, 2011, pp. 312–314.

[55] C. Niclass, M. Sergio, and E. Charbon, "A CMOS 64x48 Single Photon Avalanche Diode Array with Event-Driven Readout," *Proc. Eur. Solid-State Circuits Conf.*, pp. 556–559, 2006.

[56] C. Niclass, C. Favi, T. Kluter, M. Gersbach, and E. Charbon, "A 128 x 128 Single-Photon Image Sensor with Column Level 10-bit Time-to-digital Converter Array," *IEEE J. Solid State Circuits*, vol. 43, no. 12, pp. 2977–2989, 2008.

[57] L. Pancheri, N. Massari, F. Borghetti, and D. Stoppa, "A 32x32 SPAD Pixel Array with Nanosecond Gating and Analog Readout," in *International Image Sensor Workshop*, 2011, p. R40.

[58] J. Ma and E. Fossum, "Quanta Image Sensor Jot with Sub 0.3e- r.m.s. Read Noise and Photon Counting Capability," *IEEE Electron Device Lett.*, vol. PP, no. 99, pp. 1–1, 2015.

[59] K. Kotani, T. Shibata, and T. Ohmi, "CMOS charge-transfer preamplifier for offset-fluctuation cancellation in low-power A/D converters," *IEEE J. Solid-State Circuits*, vol. 33, no. 5, pp. 762–769, May 1998.

[60] F. Hurter and V. C. Driffield, "Photo-chemical Investigations," *Soc. Chem. Ind.*, no. 9, 455, 1890.

[61] E. R. Fossum, "Modeling the Performance of Single-Bit and Multi-Bit Quanta Image Sensors," *IEEE J. Electron Devices Soc.*, vol. 1, no. 9, pp. 166–174, Sep. 2013.

[62] F. Remondino and D. Stoppa, Eds., *TOF Range-Imaging Cameras*, 1st ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013.

[63] G. Marconi, "Radio telegraphy," *J. Am. Inst. Electr. Eng.*, vol. 41, no. 8, pp. 561–570, Aug. 1922.

[64] D. D.-U. Li, S. Ameer-Beg, J. Arlt, D. Tyndall, R. Walker, D. R. Matthews, V. Visitkul, J. Richardson, and R. K. Henderson, "Time-domain fluorescence lifetime imaging techniques suitable for solid-state imaging sensor arrays.," *Sensors (Basel).*, vol. 12, no. 5, pp. 5650–69, Jan. 2012.

[65] D. Stoppa and L. Pancheri, "D. Stoppa, L. Pancheri, M. Perenzoni - Image Sensors - London, 20th March 2012 7.1," in *Image Sensors 2012 - Workshop 1*, 2012, no. March, pp. 1–6.

[66] D. Stoppa, L. Pancheri, M. Scandiuzzo, L. Gonzo, G.-F. Dalla Betta, and A. Simoni, "A CMOS 3-D Imager Based on Single Photon Avalanche Diode," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 54, no. 1, pp. 4–12, Jan. 2007.
[67] E. A. G. Webster, "Single-Photon Avalanche Diode Theory, Simulation, and High-Performance CMOS Integration," The University of Edinburgh, 2013.
[68] E. Charbon, "Single-photon imaging in complementary metal oxide semiconductor processes," *Phil Trans R Soc A*, vol. 372, no. 2012, Feb. 2014.
[69] E. A. G. Webster, L. A. Grant, and R. K. Henderson, "Transient Single-Photon Avalanche Diode Operation, Minority Carrier Effects, and Bipolar Latch Up," *IEEE Trans. Electron Devices*, vol. 60, no. 3, pp. 1188–1194, Mar. 2013.
[70] S. Cova, A. Longoni, and A. Andreoni, "Towards picosecond resolution with single-photon avalanche diodes," *Rev. Sci. Instrum.*, vol. 52, no. 3, pp. 408–412, 1981.
[71] A. Rochas, "Single Photon Avalanche Diodes in CMOS Technology," EPFL, 2003.
[72] A. Rochas, M. Gosch, A. Serov, P. A. Besse, R. S. Popovic, T. Lasser, and R. Rigler, "First fully integrated 2-D array of single-photon detectors in standard CMOS technology," *IEEE Photonics Technol. Lett.*, vol. 15, no. 7, pp. 963–965, Jul. 2003.
[73] E. R. Fossum, "Active pixel sensors: are CCDs dinosaurs?," in *IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology*, 1993, pp. 2–14.
[74] D. Tyndall, B. Rae, D. Li, J. Richardson, J. Arlt, and R. Henderson, "A 100Mphoton/s time-resolved mini-silicon photomultiplier with on-chip fluorescence lifetime estimation in 0.13μm CMOS imaging technology," in *2012 IEEE International Solid-State Circuits Conference*, 2012, pp. 122–124.
[75] Y. Maruyama, J. Blacksberg, and E. Charbon, "A 1024x8, 700-ps Time-Gated SPAD Line Sensor for Planetary Surface Exploration With Laser Raman Spectroscopy and LIBS," *IEEE J. Solid-State Circuits*, vol. 49, no. 1, pp. 179–189, Jan. 2014.
[76] N. Krstajić, J. Levitt, S. Poland, S. Ameer-Beg, and R. Henderson, "256 × 2 SPAD line sensor for time resolved fluorescence spectroscopy.," *Opt. Express*, vol. 23, no. 5, pp. 5653–69, Mar. 2015.
[77] J. Arlt, D. Tyndall, B. R. Rae, D. D.-U. Li, J. A. Richardson, and R. K. Henderson, "A study of pile-up in integrated time-correlated single photon counting systems.," *Rev. Sci. Instrum.*, vol. 84, no. 10, p. 103105, Oct. 2013.
[78] C. Niclass, K. Ito, M. Soga, H. Matsubara, I. Aoyagi, S. Kato, and M. Kagami, "Design and characterization of a 256 x 64-pixel single-photon imager in CMOS for a MEMS-based laser scanning time-of-flight sensor.," *Opt. Express*, vol. 20, no. 11, pp. 11863–81, May 2012.
[79] E. Venialgo, E. Charbon, and S. Mandai, "MD-SiPM PET Detector Module Design," in *IEEE Nuclear Science Symposium*, 2014.
[80] E. Gros-Daillon, L. Maingault, L. André, V. Reboud, L. Verger, E. Charbon, C. Bruschini, C. Veerappan, D. Stoppa, N. Massari, M. Perenzoni, L. H. C. Braga, L. Gasparini, R. K. Henderson, R. Walker, S. East, L. Grant, B. Jatekos, E. Lorincz, F. Ujhelyi, G. Erdei, P. Major, Z. Papp, and G. Nemeth, "First characterization of the SPADnet sensor: a digital silicon photomultiplier for PET applications," *J. Instrum.*, vol. 8, no. 12, pp. C12026–C12026, Dec. 2013.
[81] L. H. C. Braga, L. Pancheri, L. Gasparini, M. Perenzoni, R. Walker, R. K. Henderson, and D. Stoppa, "A CMOS mini-SiPM detector with in-pixel data compression for PET applications," in *2011 IEEE Nuclear Science Symposium Conference Record*, 2011, pp. 548–552.
[82] L. H. C. Braga, L. Gasparini, L. Grant, R. K. Henderson, N. Massari, M. Perenzoni, D. Stoppa, and R. Walker, "A Fully Digital 8x16 SiPM Array for PET Applications With Per-Pixel TDCs and Real-Time Energy Output," *IEEE J. Solid-State Circuits*, vol. PP, no. 99, pp. 1–14, 2014.
[83] T. Frach, G. Prescher, C. Degenhardt, R. de Gruyter, A. Schmitz, and R. Ballizany, "The digital silicon photomultiplier — Principle of operation and intrinsic detector performance," in *2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*, 2009, pp. 1959–1965.
[84] A. Carimatto, S. Mandai, E. Venialgo, T. Gong, G. Borghi, D. R. Schaart, and E. Charbon, "11.4 A 67,392-SPAD PVTB-compensated multi-channel digital SiPM with 432 column-parallel 48ps 17b TDCs for endoscopic time-of-flight PET," in *IEEE International Solid-State Circuits Conference - ISSCC Digest of Technical Papers*, 2015, pp. 1–3.

[85] T. Frach, G. Prescher, C. Degenhardt, and B. Zwaans, "The digital silicon photomultiplier — System architecture and performance evaluation," in *IEEE Nuclear Science Symposuim &Medical Imaging Conference*, 2010, pp. 1722–1727.
[86] A. Spinelli, L. M. Davis, and H. Dautet, "Actively quenched single-photon avalanche diode for high repetition rate time-gated photon counting," *Rev. Sci. Instrum.*, vol. 67, no. 1, p. 55, 1996.
[87] F. Villa, D. Bronzi, Y. Zou, C. Scarcella, G. Boso, S. Tisa, A. Tosi, F. Zappa, D. Durini, S. Weyers, U. Paschen, and W. Brockherde, "CMOS SPADs with up to 500 μm diameter and 55% detection efficiency at 420 nm," *J. Mod. Opt.*, vol. 61, no. 2, pp. 102–115, Jan. 2014.
[88] D. Stoppa, D. Mosconi, L. Pancheri, and L. Gonzo, "Single-Photon Avalanche Diode CMOS Sensor for Time-Resolved Fluorescence Measurements," *IEEE Sens. J.*, vol. 9, no. 9, pp. 1084–1090, Sep. 2009.
[89] A. Eisele, R. Henderson, B. Schmidtke, T. Funk, L. Grant, J. Richardson, and W. Freude, "185 MHz Count Rate , 139 dB Dynamic Range Single-Photon Avalanche Diode with Active Quenching Circuit in 130 nm CMOS Technology," *Int. Image Sensors Work.*, pp. 6–8, 2011.
[90] R. G. Brown, R. Jones, J. G. Rarity, and K. D. Ridley, "Characterization of silicon avalanche photodiodes for photon correlation measurements. 2: Active quenching.," *Appl. Opt.*, vol. 26, no. 12, pp. 2383–9, Jun. 1987.
[91] E. Fisher, I. Underwood, and R. Henderson, "A Reconfigurable Single-Photon-Counting Integrating Receiver for Optical Communications," *IEEE J. Solid-State Circuits*, vol. 48, no. 7, pp. 1638–1650, Jul. 2013.
[92] J. A. Richardson, E. A. G. Webster, L. A. Grant, and R. K. Henderson, "Scaleable Single-Photon Avalanche Diode Structures in Nanometer CMOS Technology," *IEEE Trans. Electron Devices*, vol. 58, no. 7, pp. 2028–2035, Jul. 2011.
[93] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, "Avalanche photodiodes and quenching circuits for single-photon detection.," *Appl. Opt.*, vol. 35, no. 12, pp. 1956–76, Apr. 1996.
[94] D. Stoppa, L. Pancheri, M. Scandiuzzo, A. Simoni, L. Viarani, and G.-F. D. Betta, "A CMOS Sensor based on Single Photon Avalanche Diode for Distance Measurement Applications," *2005 IEEE Instrumentationand Meas. Technol. Conf. Proc.*, vol. 2, 2005.
[95] R. J. Walker, E. A. G. Webster, J. Li, N. Massari, and R. K. Henderson, "High fill factor digital Silicon Photomultiplier structures in 130nm CMOS imaging technology," in *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)*, 2012, pp. 1945–1948.
[96] A. Eisele, "Millimeter-Precision Laser Rangefinder Using a Low-Cost Photon Counter," Karlsruhe University, 2014.
[97] E. A. G. Webster, J. A. Richardson, L. A. Grant, D. Renshaw, and R. K. Henderson, "An Infra-Red Sensitive , Low Noise , Single- Photon Avalanche Diode in 90nm CMOS," in *2011 International Image Sensors Workshop*, 2011.
[98] J. A. Richardson, "Time Resolved Single Photon Imaging in Nanometer Scale CMOS Technology," no. May, 2010.
[99] R. H. Haitz, A. Goetzberger, R. M. Scarlett, and W. Shockley, "Avalanche Effects in Silicon p—n Junctions. I. Localized Photomultiplication Studies on Microplasmas," *J. Appl. Phys.*, vol. 34, no. 6, p. 1581, 1963.
[100] W. J. . Kindt, "A novel avalanche photodiode array," in *Proceedings of 1994 IEEE Nuclear Science Symposium - NSS'94*, 1994, vol. 1, pp. 164–167.
[101] C. Niclass, M. Gersbach, R. Henderson, L. Grant, and E. Charbon, "A Single Photon Avalanche Diode Implemented in 130-nm CMOS Technology," *IEEE J. Sel. Top. Quantum Electron.*, vol. 13, no. 4, pp. 863–869, 2007.
[102] H. Finkelstein, M. J. Hsu, and S. C. Esener, "STI-Bounded Single-Photon Avalanche Diode in a Deep-Submicrometer CMOS Technology," *IEEE Electron Device Lett.*, vol. 27, no. 11, pp. 887–889, 2006.
[103] H. Finkelstein, M. J. Hsu, S. Zlatanovic, and S. Esener, "Performance trade-offs in single-photon avalanche diode miniaturization.," *Rev. Sci. Instrum.*, vol. 78, no. 10, p. 103103, 2007.
[104] M. J. Hsu, H. Finkelstein, and S. C. Esener, "A CMOS STI-Bound Single-Photon Avalanche Diode With 27-ps Timing Resolution and a Reduced Diffusion Tail," *Electron Device Lett. IEEE*, vol. 30, no. 6, pp. 641–643, 2009.
[105] R. K. Henderson, E. A. G. Webster, R. Walker, J. A. Richardson, and L. A. Grant, "A 3×3, 5μm pitch, 3-transistor single photon avalanche diode array with integrated 11V bias generation in 90nm CMOS technology," in *2010 International Electron Devices Meeting*, 2010, pp. 14.2.1–14.2.4.

[106] J. A. Richardson, "Time Resolved Single Photon Imaging in Nanometer Scale CMOS Technology," 2010.
[107] J. A. Richardson, E. A. G. Webster, L. A. Grant, and R. K. Henderson, "Scaling trends of single-photon avalanche diode arrays in nanometer CMOS technology," in *Proceedings of SPIE*, 2011, vol. 8033, no. 1.
[108] R. L. Nicol and A. J. Holmes, "Private Conversation." 2012.
[109] E. A. G. Webster, J. A. Richardson, L. A. Grant, and R. K. Henderson, "Single-Photon Avalanche Diodes in 90nm CMOS imaging technology with sub-1Hz Median Dark Count Rate," in *2011 International Image Sensors Workshop*, 2011, pp. 7–10.
[110] C. Veerappan and E. Charbon, "A substrate isolated CMOS SPAD enabling wide spectral response and low electrical crosstalk," *IEEE J. Sel. Top. Quantum Electron.*, vol. PP, no. 99, pp. 1–1, 2014.
[111] M. Ghioni, A. Gulinatti, I. Rech, F. Zappa, and S. Cova, "Progress in Silicon Single-Photon Avalanche Diodes," *IEEE J. Sel. Top. Quantum Electron.*, vol. 13, no. 4, pp. 852–862, 2007.
[112] A. Lacaita, S. Cova, M. Ghioni, and F. Zappa, "Single-photon avalanche diode with ultrafast pulse response free from slow tails," *IEEE Electron Device Lett.*, vol. 14, no. 7, pp. 360–362, 1993.
[113] M. J. Hsu, S. C. Esener, and H. Finkelstein, "A CMOS STI-Bound Single-Photon Avalanche Diode With 27-ps Timing Resolution and a Reduced Diffusion Tail," *IEEE Electron Device Lett.*, vol. 30, no. 6, pp. 641–643, Jun. 2009.
[114] J. M. Pavia, M. Wolf, and E. Charbon, "Measurement and modeling of microlenses fabricated on single-photon avalanche diode arrays for fill factor recovery.," *Opt. Express*, vol. 22, no. 4, pp. 4202–13, Feb. 2014.
[115] C. Dege, G. Prescher, T. Frach, A. Thon, R. de Gruyter, A. Schmitz, and R. Ballizany, "The digital Silicon Photomultiplier — A novel sensor for the detection of scintillation light," in *2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*, 2009, pp. 2383–2386.
[116] G. Zhang, C. Yu, C. Zhu, and L. Liu, "Feasibility study of Multi-Pixel Photon Counter serving as the detector in digital optical communications," *Opt. - Int. J. Light Electron Opt.*, vol. 124, no. 22, pp. 5781–5786, 2013.
[117] "Hamamatsu Photonics MPPC S12572-010C," 2014. [Online]. Available: http://www.hamamatsu.com/eu/en/product/category/3100/4004/4113/S12572-010C/index.html. [Accessed: 02-May-2014].
[118] L. H. C. Braga, L. Pancheri, L. Gasparini, R. K. Henderson, and D. Stoppa, "A mini-SiPM array for PET detectors implemented in 0.35-μm HV CMOS technology," in *PhD Research in Microelectronics and Electronics PRIME*, 2011, pp. 181–184.
[119] D. Tyndall, B. R. Rae, D. D.-U. Li, J. Arlt, A. Johnston, J. A. Richardson, and R. K. Henderson, "A high-throughput time-resolved mini-silicon photomultiplier with embedded fluorescence lifetime estimation in 0.13 μm CMOS.," *IEEE Trans. Biomed. Circuits Syst.*, vol. 6, no. 6, pp. 562–70, Dec. 2012.
[120] D. Tyndall, "A CMOS System for High Throughput Fluorescence Lifetime Sensing using Time Correlated Single Photon Counting," The University of Edinburgh, 2013.
[121] C. Niclass, A. Rochas, P.-A. Besse, and E. Charbon, "Design and Characterization of a CMOS {3-D} Image Sensor Based on Single Photon Avalanche Diodes," *IEEE J. Solid-State Circuits*, vol. 40, no. 9, pp. 1847–1854, 2005.
[122] C. Niclass, C. Favi, T. Kluter, M. Gersbach, and E. Charbon, "A 128×128 Single-Photon Imager with on-Chip Column-Level 10b Time-to-Digital Converter Array Capable of 97ps Resolution," in *2008 IEEE International SolidState Circuits Conference Digest of Technical Papers*, 2008, no. i, pp. 44–594.
[123] C. Niclass and M. Soga, "A miniature actively recharged single-photon detector free of afterpulsing effects with 6ns dead time in a 0.18μm CMOS technology," in *2010 International Electron Devices Meeting*, 2010, pp. 14.3.1–14.3.4.
[124] C. Niclass, M. Soga, and S. Kato, "A 0.18μm CMOS single-photon sensor for coaxial laser rangefinders," in *2010 IEEE Asian Solid-State Circuits Conference*, 2010, pp. 1–4.
[125] R. M. Field, S. Realov, and K. L. Shepard, "A 100-fps, Time-Correlated Single-Photon-Counting-Based Fluorescence-Lifetime Imager in 130-nm CMOS," *IEEE J. Solid State Circuits*, vol. 49, no. 4, pp. 1–14, 2014.
[126] R. M. Field and K. L. Shepard, "A 100-fps fluorescence lifetime imager in standard 0.13-μm CMOS." pp. C10–C11, 2013.

[127] B. F. Aull, A. H. Loomis, J. A. Gregory, and D. J. Young, "Geiger-mode avalanche photodiode arrays integrated with CMOS timing circuits," in *56th Annual Device Research Conference Digest (Cat. No.98TH8373)*, 1998, pp. 58–59.
[128] M. Gersbach, R. Trimananda, Y. Maruyama, M. Fishburn, D. Stoppa, J. Richardson, R. Walker, R. K. Henderson, and E. Charbon, "High frame-rate TCSPC-FLIM using a novel SPAD-based image sensor," in *Proceedings of SPIE*, 2010, vol. 7780, no. 1, p. 77801H–77801H–13.
[129] D. D.-U. Li, J. Arlt, D. Tyndall, R. Walker, J. Richardson, D. Stoppa, E. Charbon, and R. K. Henderson, "Video-rate fluorescence lifetime imaging camera with CMOS single-photon avalanche diode arrays and high-speed imaging algorithm.," *J. Biomed. Opt.*, vol. 16, no. 9, p. 096012, Sep. 2011.
[130] D.-U. Li, J. Arlt, J. Richardson, R. Walker, A. Buts, D. Stoppa, E. Charbon, and R. Henderson, "Real-time fluorescence lifetime imaging system with a 32 × 32 0.13um CMOS low dark-count single-photon avalanche diode array," *Opt. Express*, vol. 18, no. 10, pp. 10257–10269, 2010.
[131] F. Villa, B. Markovic, S. Bellisai, D. Bronzi, A. Tosi, F. Zappa, S. Tisa, D. Durini, S. Weyers, U. Paschen, and W. Brockherde, "SPAD Smart Pixel for Time-of-Flight and Time-Correlated Single-Photon Counting Measurements," *IEEE Photonics J.*, vol. 4, no. 3, pp. 795–804, Jun. 2012.
[132] D. Bronzi, F. Villa, S. Bellisai, B. Markovic, G. Boso, C. Scarcella, A. Della Frera, and A. Tosi, "CMOS SPAD pixels for indirect time-of-flight ranging," in *IEEE Photonics Conference 2012*, 2012, pp. 22–23.
[133] B. Markovic, S. Bellisai, and F. A. Villa, "15bit Time-to-Digital Converters with 0.9% DNL rms and 160ns FSR for single-photon imagers," in *2011 7th Conference on Ph.D. Research in Microelectronics and Electronics*, 2011, pp. 25–28.
[134] F. Villa, R. Lussana, D. Tamborini, D. Bronzi, B. Markovic, A. Tosi, F. Zappa, and S. Tisa, "CMOS single photon sensor with in-pixel TDC for Time-of-Flight applications," in *2013 IEEE Nordic-Mediterranean Workshop on Time-to-Digital Converters (NoMe TDC)*, 2013, pp. 1–6.
[135] I. Vornicu, R. Carmona-Galan, and A. Rodriguez-Vazquez, "A CMOS 0.18μm 64×64 single photon image sensor with in-pixel 11b time-to-digital converter," in *2014 International Semiconductor Conference (CAS)*, 2014, pp. 131–134.
[136] Y. Maruyama and E. Charbon, "A Time-Gated 128X128 CMOS SPAD Array for On-Chip Fluorescence Detection," in *2011 International Image Sensors Workshop*, 2011.
[137] M. Gersbach, R. Trimananda, Y. Maruyama, M. Fishburn, D. Stoppa, J. Richardson, R. Walker, R. K. Henderson, and E. Charbon, "High frame-rate TCSPC-FLIM using a novel SPAD-based image sensor," in *Proceedings of SPIE*, 2010, vol. 7780, no. 1, p. 77801H–77801H–13.
[138] S. Burri, F. Powolny, C. E. Bruschini, X. Michalet, F. Regazzoni, and E. Charbon, "65K pixel, 150k frames-per-second camera with global gating and micro-lenses suitable for life-time imaging," in *Proc. of SPIE Photonics Europe*, 2014.
[139] S. Burri, Y. Maruyama, X. Michalet, F. Regazzoni, C. Bruschini, and E. Charbon, "Architecture and applications of a high resolution gated SPAD image sensor," *Opt. Express*, vol. 22, no. 14, pp. 17573–89, Jul. 2014.
[140] B. F. Aull, D. R. Schuette, D. J. Young, D. M. Craig, B. J. Felton, and K. Warner, "A Study of Crosstalk in a 256x256 Photon Counting Imager Based on Silicon Geiger-Mode Avalanche Photodiodes," *IEEE Sens. J.*, vol. 15, no. 4, pp. 2123–2132, Apr. 2015.
[141] D. Stoppa, F. Borghetti, J. Richardson, R. Walker, L. Grant, R. K. Henderson, M. Gersbach, and E. Charbon, "A 32x32-pixel array with in-pixel photon counting and arrival time measurement in the analog domain," in *Proceedings of the European Solid-State Circuits Conference (ESSCIRC)*, 2009, pp. 204–207.
[142] D. Chitnis and S. Collins, "Compact readout circuits for SPAD arrays," in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, 2010, pp. 357–360.
[143] D. Chitnis and S. Collins, "A flexible compact readout circuit for SPAD arrays," in *SPIE NanoScience + Engineering*, 2010, p. 77801E–77801E–9.
[144] L. Pancheri, N. Massari, and D. Stoppa, "SPAD Image Sensor With Analog Counting Pixel for Time-Resolved Fluorescence Detection," *IEEE Trans. Electron Devices*, vol. 60, no. 10, pp. 3442–3449, Oct. 2013.
[145] E. Panina, G.-F. Dalla Betta, L. Pancheri, and D. Stoppa, "Design of CMOS Gated Analog Readout Circuits for SPAD Pixel Arrays," in *Ph.D. Research in Microelectronics and Electronics (PRIME), 2012 8th Conference on*, 2012.
[146] L. Pancheri, E. Panina, G.-F. Dalla Betta, L. Gasparini, and D. Stoppa, "Compact analog counting SPAD pixel with 1.9% PRNU and 530ps time gating," in *Proceedings of the European Solid-State Circuits Conference (ESSCIRC)*, 2013, pp. 295–298.

[147] E. Panina, L. Pancheri, G.-F. Dalla Betta, L. Gasparini, and D. Stoppa, "Compact CMOS analog readout circuit for photon counting applications," in *SPIE Optics + Optoelectronics*, 2013, p. 877305.
[148] E. Panina, L. Pancheri, G. Dalla Betta, N. Massari, and D. Stoppa, "Compact CMOS Analog Counter for SPAD Pixel Arrays," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 61, no. 4, pp. 214–218, Apr. 2014.
[149] M. Perenzoni, N. Massari, D. Perenzoni, L. Gasparini, and D. Stoppa, "A 160x120 Pixel Analog-Counting Single-Photon Imager with Sub-ns Time Gating and Self-Referenced Column-Parallel A/D Conversion for Fluorescence Lifetime Imaging," in *IEEE International Solid-State Circuits Conference - ISSCC Digest of Technical Papers*, 2015, p. 11.3.
[150] L. Parmesan, "Unpublished - Fluorescence Lifetime Centre of Mass Pixel," 2015.
[151] J. A. Richardson, L. A. Grant, and R. K. Henderson, "Low Dark Count Single-Photon Avalanche Diode Structure Compatible With Standard Nanometer Scale CMOS Technology," in *Proc. of 2009 International Image Sensor Workshop, Bergen, NORWAY,* 2009, vol. 21, no. 14, pp. 1020–1022.
[152] "Single Photon detectors for medical scanning applications." [Online]. Available: http://ec.europa.eu/digital-agenda/en/news/single-photon-detectors-medical-scanning-applications. [Accessed: 17-Jun-2015].
[153] S. K. Mendis, B. Pain, R. H. Nixon, and E. R. Fossum, "Low-light-level image sensor with on-chip signal processing," in *Optical Engineering and Photonics in Aerospace Sensing*, 1993, pp. 23–33.
[154] W. J. Marble and D. T. Comer, "Ultra low power A/D converters using an enhanced differential charge-transfer amplifier." pp. 216–219, 2000.
[155] W. J. Marble and D. T. Comer, "Analysis of the dynamic behavior of a charge-transfer amplifier," *IEEE Trans. Circuits Syst. I Fundam. Theory Appl.*, vol. 48, no. 7, pp. 793–804, Jul. 2001.
[156] W. J. Marble, K. Kotani, and C. S. Petrie, "Practical Charge-Transfer Amplifier Design Architectures for Low-Power Flash A/D Converters," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 51, no. 11, pp. 2157–2164, Nov. 2004.
[157] B. R. Rae, C. Griffin, J. McKendry, J. M. Girkin, H. X. Zhang, E. Gu, D. Renshaw, E. Charbon, M. D. Dawson, and R. K. Henderson, "CMOS driven micro-pixel LEDs integrated with single photon avalanche diodes for time resolved fluorescence measurements," *J. Phys. D. Appl. Phys.*, vol. 41, no. 9, p. 094011, May 2008.
[158] S. K. Mendis, S. E. Kemeny, R. C. Gee, B. Pain, C. O. Staller, and E. R. Fossum, "CMOS active pixel image sensors for highly integrated imaging systems," *Solid-State Circuits, IEEE J.*, vol. 32, no. 2, pp. 187–197, 1997.
[159] B. Razavi, *Principles of Data Conversion System Design*. Piscataway, NJ: IEEE Press, 1995.
[160] B. Razavi, *Monolithic Phase-Locked Loops and Clock Recovery Circuits: Theory and Design*. John Wiley & Sons, 1996.
[161] R. K. Henderson, J. E. D. Hurwitz, L. A. Grant, K. M. Findlater, and T. Lule, "Pixel-Pixel Fixed Pattern Noise in CMOS Image Sensors due to Readout Parasitic Capacitance," in *International Image Sensor Workshop*, 2005.
[162] S. K. Mendis, S. E. Kemeny, and E. R. Fossum, "A 128×128 CMOS active pixel image sensor for highly integrated imaging systems," in *Proceedings of IEEE International Electron Devices Meeting*, pp. 583–586.
[163] L. Pancheri, D. Stoppa, and G.-F. Dalla Betta, "Characterization and Modeling of Breakdown Probability in Sub-Micrometer CMOS SPADs," *IEEE J. Sel. Top. Quantum Electron.*, vol. 20, no. 6, pp. 328–335, Nov. 2014.
[164] N. Teranishi, "Required Conditions for Photon-Counting Image Sensors," *IEEE Trans. Electron Devices*, vol. 59, no. 8, pp. 2199–2205, Aug. 2012.
[165] A. J. P. Theuwissen, "Harvest Imaging - How to Measure Full Well Capacity," 2013. [Online]. Available: http://harvestimaging.com/blog/?p=1238. [Accessed: 19-Aug-2015].
[166] J. Raymond Davis and J. F. M. Walters, "Sensitometry of Photographic Emulsions and a Survey of the Characteristics of Plates and Films of American Manufacture," *Sci. Pap. Bur. Stand.*, vol. 18, no. 439, 1923.
[167] S. Nasuno, S. Wakashima, F. Kusuhara, R. Kuroda, and S. Sugawa, "A CMOS Image Sensor with 240μV/eConversion Gain, 200keFull Well Capacity and 190-1000nm Spectral Response," in *International Image Sensor Workshop*, 2015.
[168] Andor, "Andor EMCCD iXon 860 Specifications," 2015. [Online]. Available: http://www.andor.com/pdfs/specifications/Andor_iXon_860_Specifications.pdf. [Accessed: 21-Aug-2015].

[169] P. Fereyre, F. Mayer, M. Fournier, C. Buton, T. Brugière, and R. Barbier, "Electron Multiplying Device Made on a 180 nm Standard CMOS Imaging Technology," in *International Image Sensor Workshop*, 2015.
[170] T. Resetar, K. De Munck, L. Haspeslagh, P. De Moor, P. Goetschalckx, R. Puers, and C. Van Hoof, "Backside-Illuminated 4-T Pinned Avalanche Photodiode Pixel for Readout Noise-Limited Applications," in *International Image Sensor Workshop*, 2015.
[171] S. Masoodian, A. Rao, J. Ma, K. Odame, and E. R. Fossum, "A 2.5pJ/b Readout Circuit for 1000fps Single-bit Quanta Image Sensors," in *International Image Sensor Workshop*, 2015.
[172] D. Stoppa, N. Massari, L. Pancheri, M. Malfatti, M. Perenzoni, and L. Gonzo, "An 80×60 range image sensor based on 10μm 50MHz lock-in pixels in 0.18μm CMOS," in *IEEE International Solid-State Circuits Conference - ISSCC Digest of Technical Papers*, 2010, pp. 406–407.
[173] W. Kim, W. Yibing, I. Ovsiannikov, S. Lee, Y. Park, C. Chung, and E. Fossum, "A 1.5Mpixel RGBZ CMOS image sensor for simultaneous color and range image capture," in *IEEE International Solid-State Circuits Conference - ISSCC Digest of Technical Papers*, 2012, pp. 392–394.
[174] C. S. Bamji, P. O'Connor, T. Elkhatib, S. Mehta, B. Thompson, L. A. Prather, D. Snow, O. C. Akkaya, A. Daniel, A. D. Payne, T. Perry, M. Fenton, and V.-H. Chan, "A 0.13 μm CMOS System-on-Chip for a 512 × 424 Time-of-Flight Image Sensor With Multi-Frequency Photo-Demodulation up to 130 MHz and 2 GS/s ADC," *IEEE J. Solid-State Circuits*, vol. 50, no. 1, pp. 303–319, Jan. 2015.
[175] T. Vogelsang, M. Guidash, and S. Xue, "Overcoming the Full Well Capacity Limit: High Dynamic Range Imaging Using Multi-Bit Temporal Oversampling and Conditional Reset," in *International Image Sensor Workshop*, 2013.
[176] J. M. Pavia, M. Wolf, and E. Charbon, "A Dual Backside-Illuminated 800-Cell Multi-Channel Digital SiPM with 100 TDCs in 130nm 3D IC Technology," in *IEEE Nuclear Science Symposium*, 2014.
[177] C. L. Sholes, "Type Writer," US 558428 A1881.
[178] D. (Stanford R. I. Engelbart, "X-y position indicator for a display system," US3541541 A17-Nov-1970.
[179] P. (PCWorld) Cohen, "Macworld Expo Keynote Live Update | Macworld," *Macworld*, 2007. [Online]. Available: http://www.macworld.com/article/1054764/liveupdate.html. [Accessed: 28-Mar-2014].
[180] Apple, "Mac Basics: Multi-Touch Gestures," 2014. [Online]. Available: http://support.apple.com/kb/ht4721. [Accessed: 13-Mar-2014].
[181] S. Israel, "Why Apple Bought PrimeSense - Forbes," 2013. [Online]. Available: http://www.forbes.com/sites/shelisrael/2013/11/25/why-would-apple-buy-primesense/. [Accessed: 28-Mar-2014].
[182] M. Corporation, "Kinect Ads: 'You Are the Controller,'" 2010. [Online]. Available: https://www.microsoft.com/en-us/news/features/2010/oct10/10-21kinectads.aspx.
[183] M. Corporation, "Xbox.com," 2014. [Online]. Available: http://www.xbox.com/en-GB/xbox-one/get-the-facts. [Accessed: 28-Mar-2014].
[184] S. Corporation, "Sony PlayStation 4," 2014. [Online]. Available: http://uk.playstation.com/ps4/features/accessories/. [Accessed: 28-Mar-2014].
[185] SoftKinetic, "SoftKinetic Depthsense Modules." [Online]. Available: http://www.softkinetic.com/products/depthsensemodules.aspx. [Accessed: 28-Mar-2014].
[186] PMDTec, "PMDTec CamBoard Pico XS," 2014. [Online]. Available: http://www.pmdtec.com/news_media/news/pico_xs.php. [Accessed: 28-Mar-2014].
[187] "YouTube - 'HP computers are racist,'" 2010. [Online]. Available: https://www.youtube.com/watch?v=t4DT3tQqgRM. [Accessed: 04-Apr-2014].
[188] "Face-Detection Cameras: Glitches Spur Charges of Racism - TIME," *Time Magazine*, 2010. [Online]. Available: http://content.time.com/time/business/article/0,8599,1954643,00.html. [Accessed: 04-Apr-2014].
[189] "Nest Smoke Alarm," 2014. [Online]. Available: https://nest.com/uk/smoke-co-alarm/life-with-nest-protect/. [Accessed: 07-Apr-2014].
[190] "Product Recall Letter from Nest CEO," 2014. [Online]. Available: https://nest.com/letter-from-the-ceo/. [Accessed: 07-Apr-2014].
[191] Sharp, "Electronic Component Catalog," 2014. [Online]. Available: http://sharp-world.com/products/device/catalog/pdf/sharp_device201403_e.pdf. [Accessed: 08-Apr-2014].

[192] Avago, "APDS-9130, Digital proximity sensors," 2014. [Online]. Available: http://www.avagotech.com/pages/en/optical_sensors/proximity_sensors/apds-9130/. [Accessed: 08-Apr-2014].
[193] C. Bouvier and Y. Ni, "Logarithmic Image Sensor For Wide Dynamic Range Stereo Vision System," in *International Image Sensor Workshop*, 2013.
[194] L. Motion, "Leap Motion Controller," 2014. [Online]. Available: https://www.leapmotion.com/. [Accessed: 08-Apr-2014].
[195] C. Dal Mutto, P. Zanuttigh, and G. M. Cortelazzo, *Time-of-Flight Cameras and Microsoft Kinect™*. Springer, 2012.
[196] R. Ng, "Digital Light Field Photography," Stanford University, 2006.
[197] E. H. Adelson and J. Y. A. Wang, "Single lens stereo with a plenoptic camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 99–106, 1992.
[198] E. H. Adelson and J. Y. A. Wang, "A stereoscopic camera employing a single main lens," in *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1991, pp. 619–624.
[199] Raytrix, "R5 Light Field Camera Systems by Raytrix - Product Specification Sheet." [Online]. Available: http://www.raytrix.de/tl_files/downloads/R5.pdf. [Accessed: 08-Apr-2014].
[200] K. Venkataraman, D. Lelescu, J. Duparré, A. McMahon, G. Molina, P. Chatterjee, R. Mullis, and S. Nayar, "PiCam," *ACM Trans. Graph.*, vol. 32, no. 6, pp. 1–13, Nov. 2013.
[201] Nikon, "Nikon Predictive Focus Tracking System," 2008. [Online]. Available: http://www.nikon.com/about/technology/rd/core/software/caf/index.htm. [Accessed: 08-Apr-2014].
[202] R. Fontaine, "The State-of-the-Art of Mainstream CMOS Image Sensors," in *International Image Sensor Workshop*, 2015, p. 1.01.
[203] D. Stoppa, "CMOS Sensors for 3D Imaging," in *IEEE International Solid-State Circuits Conference - ISSCC Forum Presentation*, 2015, pp. 1–112.
[204] B. Buttgen and P. Seitz, "Robust Optical Time-of-Flight Range Imaging Based on Smart Pixel Structures," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 55, no. 6, pp. 1512–1525, Jul. 2008.
[205] T. Spirig, P. Seitz, and O. Vietze, "Smart CCD image sensors for optical metrology and machine vision," in *AT'95: Advanced Technologies Intelligent Vision*, 1995, pp. 11–14.
[206] T. Spirig, P. Seitz, O. Vietze, and F. Heitger, "The lock-in CCD-two-dimensional synchronous detection of light," *IEEE J. Quantum Electron.*, vol. 31, no. 9, pp. 1705–1708, 1995.
[207] R. Lange and P. Seitz, "Solid-state time-of-flight range camera," *IEEE J. Quantum Electron.*, vol. 37, no. 3, pp. 390–397, Mar. 2001.
[208] R. Schwarte, Z. Xu, H. Heinol, J. Olk, R. Klein, B. Buxbaum, H. Fischer, and J. Schulte, "A new electro-optical mixing and correlating sensor: Facilities and Applications of the Photonic Mixer Device (PMD)," in *SPIE Optics + Optoelectronics*, 1997, pp. 245–253.
[209] T. Möller, H. Kraft, J. Fre, M. Albrecht, and R. Lange, "Robust 3D Measurement with PMD Sensors," in *Range Imaging Day*, 2005.
[210] B. Buttgen, T. Oggier, R. Kaufmann, P. Seitz, and N. Blanc, "Demonstration of a novel drift field pixel structure for the demodulation of modulated light waves with application in three-dimensional image capture," in *Electronic Imaging 2004*, 2004, pp. 9–20.
[211] D. Van Nieuwenhove, W. van der Tempel, and M. Kuijk, "Novel standard detector using majority current for guiding photo-generated electrons towards detecting junctions," in *Symposium IEEE/LEOS Benelux Chapter*, 2005, pp. 229–232.
[212] Q. D. Hossain, G. Dalla Betta, L. Pancheri, and D. Stoppa, "A 3D image sensor based on current assisted photonic mixing demodulator in 0.18 μm CMOS technology," in *Ph.D. Research in Microelectronics and Electronics (PRIME), 2010 Conference on*, 2010, pp. 1–4.
[213] G.-F. Dalla Betta, S. Donati, Q. D. Hossain, G. Martini, L. Pancheri, D. Stoppa, and G. Verzellesi, "TOF-range image sensor in 0.18μm CMOS technology based on current assisted photonic demodulators," in *Lasers and Electro-Optics (CLEO), 2011 Conference on*, 2011, pp. 1–2.
[214] D. Stoppa, N. Massari, L. Pancheri, M. Malfatti, M. Perenzoni, and L. Gonzo, "A Range Image Sensor Based on 10um Lock-In Pixels in 0.18um CMOS Imaging Technology," *IEEE J. Solid-State Circuits*, vol. 46, no. 1, pp. 248–258.
[215] S. Kawahito, Z. Li, and K. Yasutomi, "A CMOS Image Sensor with Draining Only Modulation Pixels for Sub-Nanosecond Time-Resolved Imaging," in *International Image Sensors Workshop*, 2011.

[216] K. Yasutomi, T. Usui, S.-M. Han, M. Kodama, T. Takasawa, K. Kagawa, and S. Kawahito, "A Time-of-Flight Image Sensor with Sub-mm Resolution Using Draining Only Modulation Pixels," in *International Image Sensor Workshop*, 2013.

[217] R. Jeremias, W. Brockherde, G. Doemens, B. Hosticka, L. Listl, and P. Mengel, "A CMOS photosensor array for 3D imaging using pulsed laser," in *2001 IEEE International Solid-State Circuits Conference. Digest of Technical Papers. ISSCC (Cat. No.01CH37177)*, 2001, pp. 252–253,.

[218] O. Elkhalili, O. M. Schrey, W. Ulfig, W. Brockherde, B. J. Hosticka, P. Mengel, and L. Listl, "A 64n Pixel 3-D CMOS Time Of Flight Image Sensor for Car Safety Applications," in *2006 Proceedings of the 32nd European Solid-State Circuits Conference*, 2006, pp. 568–571.

[219] D. Stoppa and L. Pancheri, "D. Stoppa, L. Pancheri, M. Perenzoni - Image Sensors - London, 20th March 2012 7.1 Section 6," in *Image Sensors 2012 - Workshop 1*, 2012, no. March, pp. 1–19.

[220] G. Zach and H. Zimmermann, "A 2×32 range-finding sensor array with pixel-inherent suppression of ambient light up to 120klx," in *IEEE International Solid-State Circuits Conference - ISSCC Digest of Technical Papers*, 2009, pp. 352–353,353a.

[221] G. Zach and H. Zimmermann, "Time-of-flight distance sensor with MOSFET lowpass filter as key element for suppression of ambient light," *Electron. Lett.*, vol. 46, no. 2, p. 138, 2010.

[222] G. Zach, M. Davidovic, and H. Zimmermann, "A 16x16 Pixel Distance Sensor With In-Pixel Circuitry That Tolerates 150 klx of Ambient Light," *IEEE J. Solid-State Circuits*, vol. 45, no. 7, pp. 1345–1353, Jul. 2010.

[223] M. Davidovic, G. Zach, K. Schneider-Hornstein, and H. Zimmermann, "TOF range finding sensor in 90nm CMOS capable of suppressing 180 klx ambient light," in *2010 IEEE Sensors*, 2010, pp. 2413–2416.

[224] M. Perenzoni, N. Massari, D. Stoppa, L. Pancheri, M. Malfatti, and L. Gonzo, "A 160×120-pixels range camera with on-pixel correlated double sampling and nonuniformity correction in 29.1μm pitch," in *2010 Proceedings of ESSCIRC*, 2010, pp. 294–297.

[225] M. Perenzoni, N. Massari, D. Stoppa, L. Pancheri, M. Malfatti, and L. Gonzo, "A 160 x 120-Pixels Range Camera With In-Pixel Correlated Double Sampling and Fixed-Pattern Noise Correction," *IEEE J. Solid-State Circuits*, vol. 46, no. 7, pp. 1672–1681, Jul. 2011.

[226] A. J. P. Theuwissen, "Small is Beautiful ?! Yes, but not for pixels of solid-state imagers.," in *Fraunhofer CMOS Image Sensor Workshop*, 2002.

To Helena.

Thanks for everything. With all my love,

Neale